

Beyond Bell's Theorem: Correlation Scenarios

Tobias Fritz

ABSTRACT. Bell's Theorem witnesses that the predictions of quantum theory cannot be reproduced by theories of local hidden variables in which observers can choose their measurements independently of the source. Working out an idea of Branciard, Rosset, Gisin and Pironio, we consider scenarios which feature several sources, but no choice of measurement for the observers. Every Bell scenario can be mapped into such a *correlation scenario*, and Bell's Theorem then discards those local hidden variable theories in which the sources are independent. However, most correlation scenarios do not arise from Bell scenarios, and we describe examples of (quantum) nonlocality in some of these scenarios, while posing many open problems along the way. Some of our scenarios have been considered before by mathematicians in the context of causal inference.

1. Introduction

Main ideas. Bell's Theorem [Bel64, Shi04] shows that quantum phenomena cannot be modelled correctly by a theory satisfying the following natural assumptions:

- (I) **Realism:** Any physical system can be described in terms of a probabilistic mixture of states (=hidden variable values). Composite systems are described by a joint probability distribution over the state spaces of its component systems.
- (II) **Locality:** Physical systems have spatial components which can be described independently. They do not interact across spacelike separated events.
- (III) **Free will:** The parties in a Bell scenario have genuine randomness available which is independent of their environment. This is also known as λ -*independence* [BY08] and as *measurement independence* [Hal10].

Standard quantum theory fails (I) due to the way that joint systems are described. It is irrelevant whether (III) holds in quantum theory, since (III) is only used in combination with (I) and (II) in the derivation of the Bell inequalities, which are found to have quantum violations.

In this paper, we are concerned with assumption (III). More precisely, we are actually *not* concerned with (III), since we aim to replace it with a different property:

Key words and phrases. Bell's Theorem; quantum nonlocality; inference of common ancestors.

Acknowledgements. The author would like to thank many independent sources for fruitful discussions, most of which were local and realistic, and the EU STREP QCS for financial support. Furthermore, Cyril Branciard, Lluís Masanes, Markus Müller, Nicolas Gisin, Nihat Ay and Rodrigo Gallego have provided crucial feedback on a draft version of this paper which greatly helped in improving accuracy and presentation.

- (III') Independence of sources [BGP10]: if an experiment contains several sources¹, then the theory describes these sources as independent. This means that the joint distribution of hidden variables is a product distribution.

Our observation is that (III) becomes obsolete when assuming (III'), so that one obtains:

Bell's Theorem, new version. *Quantum phenomena cannot be modelled correctly by a theory satisfying (I), (II), (III').*

Branciard, Rosset, Gisin and Pironio already briefly considered scenarios in which each party has only one measurement setting [BRGP12, Sec. V/VI]. These are a natural continuation of their earlier work [BGP10] which combined (III') with (III). Here, we build on their idea and set up a formal framework for multi-source “correlation scenarios” in which each party has only one measurement setting available and derive more results within that framework. There are several advantages to this over the standard approach based on (III):

- One of the main goals of the hidden variable program was to resurrect a deterministic worldview [EPR35]. However, as has also been observed by 't Hooft [’H07] and probably others, determinism is at variance with (III) *even without Bell's Theorem* since genuine randomness cannot be created in a deterministic world. This tension between determinism and free will has been known to philosophers long before and led them to seek definitions of human free will compatible with determinism [McK04].
- Free will is an observer-centric notion which, depending on the theory, may require the observer to live outside that part of the universe described by the theory. In contrast, the property (III') concerns only observer-independent physical systems and has clear physical meaning. our formalism is best viewed as devoid of any conscious agents.
- Bell's Theorem is often presented as a statement about theories satisfying realism (I) and locality (II) only. (III) is then tacitly assumed without explicit mention, either because one has failed to notice it as an additional and crucial assumption, or because it may be incorrectly regarded as self-evident. In contrast, (III') is more easily understood to be a non-trivial assumption.
- There has been speculation on the relation between quantum mechanics and free will. Our approach elucidates that this discussion is irrelevant to Bell's Theorem (as is well-known to experts, but possibly not to those just learning about Bell's Theorem and assumption (III)).

Moreover, our formalism allows the consideration of (quantum) correlations which have no analog in standard Bell scenarios and are genuinely new; see Theorems 2.16 and 2.21. Our current results are not sufficient to tell what the meaning or relevance of such new kinds of correlations might be; ultimately, we hope for the development of quantum information protocols utilizing them in ways similar to those taking advantage of quantum correlations in standard Bell scenarios, e.g. quantum key distribution [Eke91] or certified randomness generation [PAM⁺10]. Another interesting direction might be to consider analogs of the amplification of free will [CR12] for the amplification of independence of sources.

Inference of common ancestors. Some of the mathematical problems we are going to discuss in this paper have been considered before in a totally different context. There is work by

¹It is not perfectly clear to us what “source” actually means. One possible definition of source might be that it is a physical system which is, in the *quantum-theoretical* description, independent of its environment: the total initial state should be the tensor product of the system state and an environment state.

Steudel and Ay [SA10] on the *inference of common ancestors*, which concerns question such as this: given three different languages, under which conditions can one derive the existence of a common antecedent language which influenced all three? Or, given the joint distribution of the prevalence of some diseases in a population, under which conditions can one conclude the existence of a certain preexisting quantity or property (like a genetic defect or a specific diet) having some influence on the occurrence of all the diseases considered? This is the question of existence of a *common ancestor* in a Bayesian network model [Pea09]. A variable in a Bayesian network typically has many ancestors, including itself. One then considers models of the given joint distribution of the observed variables in terms of Bayesian networks, in which each observed variable corresponds to a node, the other nodes represent unobserved variables, and each edge represents a causal link. Then the question is whether one can find such a model without a node which is an ancestor of all the observed variables, or whether such a Bayesian network model necessarily requires such a common ancestor.

For the special case of three observed variables a, b, c , the very general results of [SA10] show that when the single-variable Shannon entropies $H(a)$, $H(b)$, $H(c)$ and the joint entropy $H(abc)$ satisfy the inequality

$$H(a) + H(b) + H(c) > 2H(abc), \quad (1.1)$$

then the existence of a common ancestor is necessary. In our example: if the vocabulary of three languages is correlated in such a way that the entropy of the joint distribution is so low that the inequality holds, then there needs to be a common precursor having influenced all three.

We will see that the inference of common ancestors is a special case of our formalism. A byproduct of our results will be an inequality similar to but strictly better than (1.1), for the very particular case of three variables; see (2.14).

Directions of future research. We hope that our ideas will spur new developments in several directions:

- Further study of classical, quantum and generalized correlations in correlation scenarios. The wealth of open problems we present shows that our results are nothing but a first step towards an understanding of correlation scenarios.
- What are the philosophical implications of our results? How do (III) and (III') compare from a philosophy of science perspective?
- Could our correlation scenarios have any relevance for applications like quantum key distribution?

A further generalization of correlation scenarios to scenarios with arbitrary causal structure will be considered in [FS12]. Correlation scenarios are a natural intermediate step between Bell scenarios and the arbitrary causal structure of [FS12].

Organization of this paper. The interested reader should start with the next subsection on terminology and notation, for otherwise the main text will not be comprehensible. The subsequent main part of the paper in Sections 2 and 3 can be read in a linear way. Section 2 contains the most important material, namely the conceptual discussion and the examples we have considered so far. Those who do not care too much about abstract generalities may stop reading at any point at which they start losing interest. In particular, reading Section 3, which contains an initial sketch of how an abstract approach to our formalism could look like, is not required for understanding the main ideas. It is supposed to be an attempt at laying the formal basis for future work on the subject.

Due to the high amount of technical detail required for completely rigorous proofs, we restrict ourselves in several cases to the presentation of proof sketches. We hope that these make it clear

how completely rigorous proofs can be constructed. In cases where a general rigorous proof or definition involves measure theory, the main text provides the proof or definition for the case of discrete hidden variables; Appendix A then treats the general case of hidden variables defined on arbitrary probability spaces.

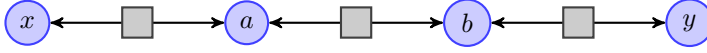
Since the subject of this paper is relatively new, many questions remain open. In the main text, we mention a wealth of open problems of various difficulties. We warn the reader that trying to solve them can be quite frustrating; our own experience has been that the intuition we have developed for standard Bell scenarios is sometimes more of a hindrance than an asset. Many of our initially promising ideas have turned out to be misconceived. Those that have eventually worked are based on very different concepts ranging from entropic inequalities (Lemma 2.14) via Hardy-type paradoxes (Theorem 2.21) to Choquet’s Theorem (see A.6). Nevertheless, we hope that our formalism will develop into an alternative approach to the study of nonlocality and will continue to be studied not only from our mathematical point of view, but also from both the information processing and the philosophical perspective. For example, the recent “PBR Theorem” [PBR11] also considers hidden variable theories satisfying (III’) and a comparison to our approach may be interesting.

Finally, Appendix A contains measure-theoretical details concerning the consideration of non-discrete hidden variables. In the main text, all our definitions and proofs are rigorous only for the case of discrete hidden variables; without exception, the same ideas work in the general case, but the technicalities required are so much more laborious and obscure that we relegate them to the appendix.

A follow-up paper [FS12] will present an even more general formalism for device-independent physics in terms of hidden Bayesian networks. It will comprise not only standard Bell scenarios and the formalism we introduce here, but also other scenarios like Popescu’s “hidden” nonlocality [Pop95]. It will be conceptually similar to hidden Markov models [LA09].

Terminology and notation. From now on, we will restrain from using the misleading term *nonlocality* and related terms like *local correlations*. It is misleading terminology insofar as it suggests that nonlocal interactions would be the only way to escape the conclusion of Bell’s Theorem; however this is far from correct, since locality is only one of the assumptions (I), (II), (III). Moreover, despite the experimental verification of the existence of quantum “nonlocality” [AGR81], all known fundamental interactions in physics are of a local nature [CG07, Haa96, Jac96]; see also [Zeh06]. Consequently, we will rather speak of *classical correlations* in analogy with the commonly used term *quantum correlations*. We will use these notions both in the context of standard Bell scenarios as well as in our new *correlation scenarios*.

In the context of our correlation scenarios, we use typewriter-font uppercase letters A, B, C, ... to enumerate the measurements. Equivalently, one may think of these as observers or parties: since each observer or party gets assigned a fixed measurements which they conduct in each run of the experiment, this is the same. The corresponding measurement outcomes are denoted by lowercase letters a, b, c, \dots . We denote the joint probability distribution of outcomes of, for example, the joint measurement (A, Y) by $p(a, y)$. This constitutes extensive abuse of notation as it makes expressions like $p(97, -2)$ ambiguous: does this refer to the distribution $p(a, y)$ or to another one like $p(w, z)$? Notwithstanding, we use this notation here in order to keep clutter to a minimum, while making sure that it does not lead to ambiguous expressions. We also keep the order of the variables arbitrary: for example, $p(x, a, b, y)$ stands for the same distribution as $p(a, b, x, y)$, and the one we use depends on which one is more natural in that particular context. Moreover, notation like $p(a, b, x, y)$ makes sense, strictly speaking, only when all variables are discrete; while we do

FIGURE 1. The correlation scenario P_4 .

assume that all measurements have only a finite number of possible outcomes, we do not make any discreteness assumption on the hidden variables; see Appendix A.

Necessary background. Any reader looking at this paper will probably already have the necessary understanding of Bell’s Theorem [Bel64, Shi04]. Moreover, we also need to assume good familiarity with the notions of (conditional) independence of random variables and conditioning of probabilities. A basic knowledge of the terminology of graphs and hypergraphs is required for Section 3. Some background in Bayesian networks [KF09, Pea09] will be of advantage in order to understand the connection to [SA10]. Reading Appendix A is not possible without some grasp of measure-theoretical probability theory and related subjects.

2. Examples of correlation scenarios

In this section, we introduce correlation scenarios by way of example. Using the appropriate dictionary from the standard framework into our formalism, we show how to translate any ordinary Bell scenario as well as the “bilocal” scenarios introduced in [BGP10] into a scenario without free will.

We also present the first examples of correlation scenarios, some of which have been considered in [BRGP12] and some of which are new. Obtaining concrete results about these new kinds of correlations has turned out to be difficult; until now, we have been able to do so only by relating to things we were already familiar with (standard Bell scenarios). We hope that future work will show the class of correlation scenarios, as we are going to formally define it in Section 3, to be much richer than what we begin to explore in this paper.

A first example. Let us consider an experimental setup as depicted abstractly in Figure 1. There are 4 parties X, A, B, Y (circles) arranged in a linear way such that any pair of neighboring parties shares a source (square). Each of these three sources sends out, at time t_{emit} , one physical system to each adjacent party. As in the case of ordinary Bell scenarios, these two systems are typically correlated; in the classical case, this is shared randomness, while in the quantum case, such a correlation can also be entanglement. The parties receive these systems and each party conducts, at time $t_{\text{meas}} > t_{\text{emit}}$, a fixed measurement on the system(s) they have received; in the case of A and B, who receive two systems each, this will typically be a joint measurement operating on both systems simultaneously. In each run of the experiment, the parties obtain and register outcomes x, a, b, y . If the experiment is repeated many times, the parties will notice correlations between these outcomes and determine a joint probability distribution $p(x, a, b, y)$. With the parties as vertices and the sources as edges, Figure 1 has the structure of the path graph P_4 , and therefore we will speak of the P_4 scenario. It has first been studied in [BRGP12, Sec. 5].

Ideally, the timing and the geometry of the experiment should guarantee that the leftmost source cannot causally influence b or y in the time between t_{emit} and t_{meas} . Similar causal separation should hold between any other pair of source and measurement which do not share an arrow in Figure 1. This ensures the validity of assumption (II).

Also, the sources should have been prepared in such a way that the correct quantum-mechanical description of the system will take the joint state of the sources to be a product state, and furthermore such that any correlation between them in a potential hidden variable description should be rendered very implausible. In other words, the experiment should try to guarantee that any hidden variable theory not satisfying (III') should be very unreasonable and contrived. This may be achieved, for example, by placing the sources at large spatial separation between each other and by using sources which employ different physical mechanisms. But of course, since the past light cones of the sources will always intersect, the requirement (III') can never be enforced. It will always be possible to explain all observations by, for example, a superdeterministic theory in which everything is predetermined since the beginning of the universe; compare [Bel87, Ch. 12].

As has already been noticed in [BGP10], this discussion is completely analogous to the discussion of the validity of property (III): there exist hidden variable theories, like superdeterminism, which do not allow free will and therefore evade the conclusion of Bell's Theorem. However, these are generally so contrived that one cannot regard them as scientific theories of physics. Exactly the same applies to our assumption (III') in a suitably conducted experiment.

Now we imagine that many runs of such an experiment have been conducted and we are given the joint outcome statistics $p(x, a, b, y)$. In the following, we work with the ideal case of infinite statistics, so that the outcome probabilities $p(x, a, b, y)$ are known with perfect precision.

Then, due to the causal structure of the experiment, one should find that the outcome x is independent of y , since X and Y do not connect to a common source. Similarly, x should be independent of b ; in fact, x should be independent of the pair (b, y) . Similarly, y should be independent of the pair (x, a) . Checking whether this is indeed the case amounts to a consistency check for the experiment.

More formally, these requirements mean that $p(x, a, b, y)$ should be a *correlation*:

DEFINITION 2.1. *A correlation p in the P_4 scenario is a distribution $p(x, a, b, y)$ whose marginals factorize as*

$$p(x, a, y) = p(x, a)p(y), \quad p(x, b, y) = p(x)p(b, y). \quad (2.1)$$

Any of these two equations implies $p(x, y) = p(x)p(y)$. Upon using this, one finds that (2.1) is equivalent to $p(a|x, y) = p(a|x)$ and $p(b|x, y) = p(b|y)$ for all those values of x and y for which $p(x) > 0$ and $p(y) > 0$. Upon reinterpreting x and y as settings in a bipartite Bell scenario having outcomes a and b , these are the no-signaling equations. However, conceptually, (2.1) has nothing to do with the impossibility of communication between the parties: these cannot do anything else than apply their fixed measurement in each run of the experiment, which renders the very notion of communication meaningless.

We now ask under which conditions a given correlation $p(x, a, b, y)$ is *classical*, i.e. consistent with the assumptions (I), (II), (III'). What would it mean to have such a model? Due to (I), the state of the systems sent out by each of the three sources can be described in terms of a classical random variable; we will denote these “hidden” variables by λ_{XA} , λ_{AB} , λ_{BY} , respectively, where the index specifies the source which the hidden variable models. For the precise definition of hidden variable, see A.1.

Assumption (III') now means that the joint distribution of these hidden variables is a product distribution:

$$p(\lambda_{XA}, \lambda_{AB}, \lambda_{BY}) = p(\lambda_{XA})p(\lambda_{AB})p(\lambda_{BY}).$$

A sensible hidden variable model should also satisfy *locality* (II): each outcome should be a (deterministic or probabilistic) function of the hidden variables associated to the sources it interacts with *and no others*.

If such a hidden variable model exists for the correlation p , then we call p *classical*. A more precise statement is this:

DEFINITION 2.2 ([BRGP12]). *A correlation $p(x, a, b, y)$ is classical in the P_4 scenario if and only if it can be written in the form*

$$p(x, a, b, y) = \int_{\lambda_{XA}, \lambda_{AB}, \lambda_{BY}} p(x|\lambda_{XA})p(\lambda_{XA})p(a|\lambda_{XA}, \lambda_{AB})p(\lambda_{AB})p(b|\lambda_{AB}, \lambda_{BY})p(\lambda_{BY})p(y|\lambda_{BY}) \quad (2.2)$$

for some collection of (conditional) distributions

$$p(x|\lambda_{XA}), \quad p(\lambda_{XA}), \quad p(a|\lambda_{XA}, \lambda_{AB}), \quad p(\lambda_{AB}), \quad p(b|\lambda_{AB}, \lambda_{BY}), \quad p(\lambda_{BY}), \quad p(y|\lambda_{BY}).$$

See A.2 for an explanation of what these conditional distributions mean in case that the hidden variables are not all discrete.

We take this to be a *definition* instead of a proposition or theorem because it is the first time that we have formalized the notion of classical model in a mathematically rigorous way. The representation (2.2) can be informally derived from hypotheses (I), (II), (III') as follows. Applying (I) and the definition of conditional probability gives

$$p(x, a, b, y) = \int_{\lambda_{XA}, \lambda_{AB}, \lambda_{BY}} p(x, a, b, y|\lambda_{XA}, \lambda_{AB}, \lambda_{BY})p(\lambda_{XA}, \lambda_{AB}, \lambda_{BY}).$$

By locality (II), the first factor in the integrand can be replaced by

$$p(a, b, x, y|\lambda_{XA}, \lambda_{AB}, \lambda_{BY}) = p(x|\lambda_{XA})p(a|\lambda_{XA}, \lambda_{AB})p(b|\lambda_{AB}, \lambda_{BY})p(y|\lambda_{BY})$$

while independence of sources (III') guarantees that the second factor is equal to

$$p(\lambda_{XA}, \lambda_{AB}, \lambda_{BY}) = p(\lambda_{XA})p(\lambda_{AB})p(\lambda_{BY}),$$

and then (2.2) directly follows.

REMARK 2.3. In the representation (2.2), it can be assumed without loss of generality that the four conditional distributions on the right-hand side are in fact deterministic, i.e. it can be assumed that the outcomes are functions

$$x = x(\lambda_{XA}), \quad a = a(\lambda_{XA}, \lambda_{AB}), \quad b = b(\lambda_{AB}, \lambda_{BY}), \quad y = y(\lambda_{BY}).$$

In the case of discrete hidden variables, this can be seen as follows: if, for example, a is a probabilistic function of λ_{XA} and λ_{AB} , then the computation of this function can be regarded as the *deterministic* computation taking the values $\lambda_{XA}, \lambda_{AB}$ and an additional random number $r_A \in [0, 1]$ as input, calculating $p(a|\lambda_{XA}, \lambda_{AB})$ for each outcome a , and then using r_A to determine which one of these finitely many outcomes occurs. But now we can redefine the hidden variable λ_{XA} to be the pair $\lambda'_{XA} = (\lambda_{XA}, r_A)$ which contains the information about the original λ_{XA} as well as the additional random number r_A required in the computation; the party X will then also receive this new component of λ'_{XA} , but can just ignore it. In this way, the function $a(\lambda'_{XA}, \lambda_{AB})$ has become deterministic.

Upon applying this kind of hidden variable redefinition for each party, all the outcomes become deterministic functions of the hidden variables.

This reasoning not only applies to P_4 , but in exactly the same way to any correlation scenario. We will make use of this in the proof of Theorem 2.21. See A.3 for a rigorous and general version of this argument.

It is also not difficult to define what *quantum correlations* are. Informally speaking, a quantum correlation is a correlation $p(x, a, b, y)$ which can be modelled in terms of quantum resources: a bipartite quantum state for each source together with one measurement for each party operating jointly on all the systems received by that party. The Hilbert space dimension of the quantum systems can be arbitrary and will be infinite in general. We take the definition of quantum correlation to be sufficiently obvious that we need to go into detail here; see Definition 3.16 for the technicalities.

The following theorem makes the connection to bipartite Bell scenarios. Its first part has also appeared in [BRGP12].

THEOREM 2.4. (1) *A correlation $p(a, b, x, y)$ is classical in P_4 if and only if the associated conditional distribution $p(a, b|x, y)$ is classical in the Bell scenario sense.*
 (2) *A correlation $p(a, b, x, y)$ is quantum in P_4 if and only if the associated conditional distribution $p(a, b|x, y)$ is quantum in the Bell scenario sense.*

Note that the use of conditional probabilities here, or in any other context, does not require any particular causal structure among the variables involved.

In forming $p(a, b|x, y)$, it is implicitly assumed that all outcomes for x and y have strictly positive probability; this can always be achieved by redefining the set of outcomes to consist of only those values which occur with positive probability.

Thus, we can roughly summarize our present results as follows: by Definition 2.2, a correlation $p(a, b, x, y)$ can be interpreted in a conventional bipartite Bell scenario as a no-signaling box together with a specification of input distributions $p(x)$ and $p(y)$; and the correlation is classical (resp. quantum) if and only if the associated no-signaling box is classical (resp. quantum).

PROOF OF THEOREM 2.4. (1) Suppose that p is classical. Then

$$p(a, b|x, y) = \int_{\lambda_{AB}} p(a, b|x, y, \lambda_{AB}) p(\lambda_{AB}).$$

By the assumption (2.2), upon conditioning on λ_{AB} , the variables (a, x) are independent of the variables (b, y) ; therefore, $p(a, b|x, y, \lambda_{AB}) = p(a|x, \lambda_{AB}) p(b|y, \lambda_{AB})$, and

$$p(a, b|x, y) = \int_{\lambda_{AB}} p(a|x, \lambda_{AB}) p(b|y, \lambda_{AB}) p(\lambda_{AB}). \quad (2.3)$$

This is the standard representation of the conditional probabilities obtained from local hidden variables in a bipartite Bell scenario. In particular, $p(a, b|x, y)$ will have to satisfy all Bell inequalities.

Conversely, we start from a correlation $p(a, b, x, y)$ for which $p(a, b|x, y)$ satisfies all Bell inequalities. This means in particular that there is a hidden variable λ such that

$$p(a, b|x, y) = \int_{\lambda} p(a|x, \lambda) p(b|y, \lambda) p(\lambda)$$

Defining $\lambda_{AX} = x$, $\lambda_{BY} = y$ and $\lambda_{AB} = \lambda$ now yields a hidden variable model in the P_4 correlation scenario, i.e. the right-hand side of (2.3).

(2) Suppose that $p(a, b, x, y)$ is quantum. Then one has one bipartite quantum state at each source and one quantum measurement at each party. We think of the measurement \mathbf{X} as remotely preparing, via steering depending on the outcome x , a quantum system for \mathbf{A} . In order to ease notation, we may assume, without loss of generality, the shared state to be pure and \mathbf{X} 's measurement to be projective. Furthermore, we may take \mathbf{X} 's projective measurement to be nondegenerate; going

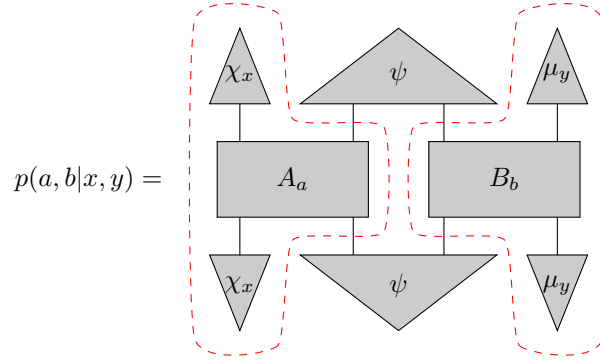
to a degenerate measurement amounts to a coarse-graining of \mathbf{X} , which preserves the quantum-mechanical realizability of $p(a, b, x, y)$. By these assumptions, the steered states for \mathbf{A} are a family $\{|\chi_x\rangle\}$ of pure states. Using the same assumptions for \mathbf{Y} , we end up with a family $\{|\mu_y\rangle\}$ of pure steered states for \mathbf{B} .

We now replace the source between \mathbf{X} and \mathbf{A} by a hidden variable defined to be $\lambda_{\mathbf{AX}} = x$; then the new measurement protocol of \mathbf{X} simply consists in announcing $\lambda_{\mathbf{AX}}$'s value as his outcome. The new protocol of \mathbf{A} consists in receiving $\lambda_{\mathbf{AX}}$, preparing the quantum state which \mathbf{X} would have steered to given the outcome $\lambda_{\mathbf{AX}}$, and then proceeding with the measurement specified in the original protocol. This replacement preserves the overall correlation $p(a, b, x, y)$. The same procedure can be applied in order to replace the source between \mathbf{Y} and \mathbf{B} by a hidden variable $\lambda_{\mathbf{BY}}$ and the measurement of \mathbf{Y} by the protocol of simply announcing $\lambda_{\mathbf{BY}}$'s value as the outcome y .

Let $\{A_a\}$ (resp. $\{B_b\}$) denote the POVM employed by \mathbf{A} (resp. \mathbf{B}). Then

$$p(a, b|x, y) = (\langle \chi_x | \otimes \langle \psi | \otimes \langle \mu_y |) (A_a \otimes B_b) (|\chi_x\rangle \otimes |\psi\rangle \otimes |\mu_y\rangle), \quad (2.4)$$

or, in graphical notation [Coe10],



Here, the dashed line indicates how to consider $A_a^x \stackrel{\text{def}}{=} \langle \chi_x | A_a | \chi_x \rangle$ as well as $B_b^y \stackrel{\text{def}}{=} \langle \mu_y | B_b | \mu_y \rangle$ as operators acting on one part of the bipartite state $|\psi\rangle$. By $\sum_a A_a = \mathbb{1}$ and normalization of $|\chi_x\rangle$, it follows that $\sum_a A_a^x = \mathbb{1}$ for all x ; similarly, $\sum_y B_b^y = \mathbb{1}$ for all y . By definition, (2.4) can then be written as

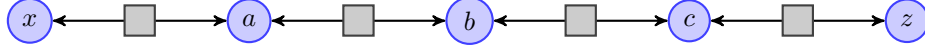
$$p(a, b|x, y) = \langle \psi | A_a^x \otimes B_b^y | \psi \rangle. \quad (2.5)$$

This is desired quantum representation of p in a bipartite Bell scenario.

Conversely, we start from a correlation $p(a, b, x, y)$ of the form (2.5). As sources between \mathbf{A} and \mathbf{X} and between \mathbf{B} and \mathbf{Y} , we again take hidden variables defined by $\lambda_{\mathbf{XA}} = x$ and $\lambda_{\mathbf{BY}} = y$; again, the protocol of \mathbf{X} and \mathbf{Y} is simply to announce the values of these variables as their outcome. Only the source between \mathbf{A} and \mathbf{B} is taken to be quantum and produces the bipartite state $|\psi\rangle$ of (2.5). The measurement protocol conducted by \mathbf{A} is similar to above: measure $\lambda_{\mathbf{XA}}$, use the result as the choice of setting for the subsequent measurement on $|\psi\rangle$, and then announce both outcomes as the total outcome. This protocol can be interpreted as measuring a single POVM given by

$$|x\rangle\langle x| \otimes A_a^x \mapsto (x, a),$$

where the left-hand side is a POVM element indexed by x and a , and the right-hand side denotes the resulting outcome announced by \mathbf{A} . The analogous POVM is measured by \mathbf{B} . By construction, this reproduces both the desired conditional distribution (2.5) and the marginal distribution $p(x, y) = p(x)p(y)$, and therefore also the whole distribution $p(x, a, b, y)$.

FIGURE 2. The correlation scenario P_5 .

□

COROLLARY 2.5. (1) *There exist non-classical quantum correlations in P_4 .*
 (2) *There exist non-quantum correlations in P_4 .*

PROOF. This follows from the existence of Bell inequality violations and no-signaling violations of Tsirelson's bound [Cir80], respectively. □

REMARK 2.6. Due to Theorem 2.4, we can regard P_4 as the analogue of a bipartite Bell scenario within our formalism. Nevertheless, there are several important differences. For one, the correlations live in completely different spaces: in a Bell scenario, one works in the space of conditional distributions $p(a, b|x, y)$, which results in the convexity of the sets of classical and quantum correlations. In contrast, in the case of P_4 , we work on the level of unconditional distributions $p(x, a, b, y)$, which contain, from the point of view of Bell scenarios, also the information about the distributions of settings $p(x)$ and $p(y)$. The sets of classical and quantum correlations in this formulation are not convex, which can be seen as follows: first, the set of classical correlations contains all the deterministic distributions $p(x, a, b, y)$ in which all measurements always produce the same outcome. Second, *any* probability distribution $p(x, a, b, y)$, and in particular every correlation, is a convex combination of deterministic ones. Third, not every correlation is classical. Thus, not every convex combination of classical correlations is a classical correlation; for that matter, most convex combinations of classical correlations are not even correlations! The same reasoning shows that the set of quantum correlations is not convex. Analogous arguments apply to any other correlation scenario in which non-classical (resp. non-quantum) correlations exist.

The scenario P_5 . We proceed to the second example of a correlation scenario. It is depicted in Figure 2. With parties as vertices and sources as edges, this is the path graph P_5 , and therefore we will speak of the P_5 scenario; the conceptual discussion we gave of the P_4 scenario applies here and to all following examples just as well. We will see that the P_5 scenario relates to the “bilocality” scenarios of Branciard, Gisin and Pironio [BGP10] (*BGP scenarios*) just as we have seen the P_4 scenario to relate standard bipartite Bell scenarios.

Given the 5-variable distribution $p(x, a, b, c, z)$, under which conditions would we expect it to arise from a configuration like Figure 2? In other words, what is the analogue of Definition 2.1? Following reasoning analogous to the P_4 case, the answer is straightforward:

DEFINITION 2.7. *A correlation p in the P_5 scenario is a distribution $p(x, a, b, c, z)$ whose marginals factorize as*

$$p(x, a, b, z) = p(x, a, b)p(z), \quad p(x, a, c, z) = p(x, a)p(c, z), \quad p(x, b, c, z) = p(x)p(b, c, z).$$

Any of these three equations implies $p(x, z) = p(x)p(z)$. Upon using this, the first and third condition can also be written as $\sum_c p(a, b, c|x, z) = \sum_c p(a, b, c|x)$ and $\sum_a p(a, b, c|x, z) = \sum_a p(a, b, c|x)$, respectively, which are formally identical to the no-signaling equations of the BGP scenario. Similarly, the second condition is then equivalent to $p(a, c|x, z) = p(a|x)p(c|z)$, which is also formally identical to a consistency constraint in the BGP scenario [CF12].

The classicality assumptions (I), (II) and (III') now yield the following characterization:

DEFINITION 2.8. A correlation $p(x, a, b, c, z)$ is classical in the P_5 scenario if and only if it can be written in the form

$$\begin{aligned} p(a, b, c, x, z) \\ = \int_{\lambda_{XA}, \lambda_{AB}, \lambda_{BC}, \lambda_{CZ}} p(x|\lambda_{XA})p(\lambda_{XA})p(a|\lambda_{XA}, \lambda_{AB})p(\lambda_{AB})p(b|\lambda_{AB}, \lambda_{BC})p(\lambda_{BC})p(c|\lambda_{BC}, \lambda_{CZ})p(\lambda_{CZ})p(z|\lambda_{CZ}) \end{aligned} \quad (2.6)$$

for some collection of (conditional) distributions

$$\begin{aligned} p(x|\lambda_{XA}), \quad p(a|\lambda_{XA}, \lambda_{AB}), \quad p(b|\lambda_{AB}, \lambda_{BC}), \quad p(c|\lambda_{BC}, \lambda_{CZ}), \quad p(z|\lambda_{CZ}). \\ p(\lambda_{XA}), \quad p(\lambda_{AB}), \quad p(\lambda_{BC}), \quad p(\lambda_{CZ}). \end{aligned}$$

As before, we regard the analogous definition of quantum correlations as straightforward and refer to 3.16 for the details.

THEOREM 2.9. (1) A correlation $p(a, b, c, x, z)$ is classical in P_5 if and only if the associated conditional distribution $p(a, b, c|x, z)$ is classical in the BGP scenario sense.
 (2) A correlation $p(a, b, c, x, z)$ is quantum in P_5 if and only if the associated conditional distribution $p(a, b, c|x, z)$ is quantum in the BGP scenario sense.

We abbreviate the proof a bit because it is completely analogous to the proof of Theorem 2.4.

PROOF. (1) Suppose that p is classical, i.e. can be written in the form (2.6). Then,

$$p(a, b, c|x, z) = \int_{\lambda_{AB}, \lambda_{BC}} p(a, b, c|x, z, \lambda_{AB}, \lambda_{BC})p(\lambda_{AB})p(\lambda_{BC}).$$

Upon conditioning on λ_{AB} and λ_{BC} , we have

$$p(a, b, c|x, z, \lambda_{AB}, \lambda_{BC}) = p(a|x, \lambda_{AB})p(b|\lambda_{AB}, \lambda_{BC})p(c|\lambda_{BC}, z),$$

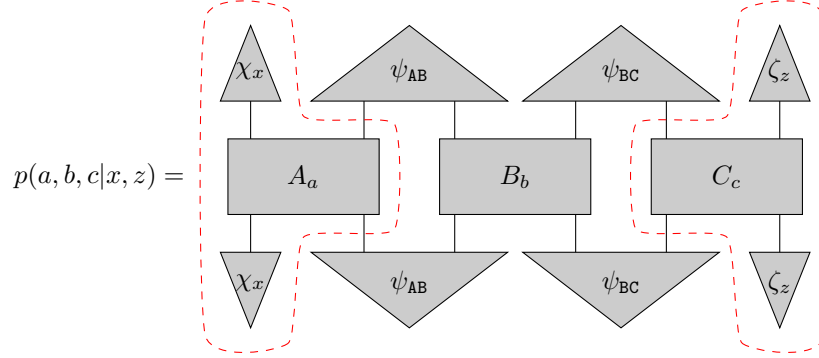
and therefore,

$$p(a, b, c|x, z) = \int_{\lambda_{AB}, \lambda_{BC}} p(a|x, \lambda_{AB})p(b|\lambda_{AB}, \lambda_{BC})p(c|\lambda_{BC}, z)p(\lambda_{AB})p(\lambda_{BC}),$$

which is the standard representation of a classical correlation in the BPG scenario [BGP10]. Conversely, upon starting from such a representation, one can again take $\lambda_{XA} = x$ and $\lambda_{BY} = y$, and (2.6) also holds.

(2) We start with a quantum correlation $p(a, b, c|x, z)$. Upon applying the same steering argument as in the proof of Theorem 2.4, we may assume, in the obvious notation,

$$\begin{aligned} p(a, b, c|x, z) = \\ (\langle \chi_x | \otimes \langle \psi_{AB} | \otimes \langle \psi_{BC} | \otimes \langle \mu_y |) (A_a \otimes B_b \otimes C_c) (|\chi_x\rangle \otimes |\psi_{AB}\rangle \otimes |\psi_{BC}\rangle \otimes |\mu_y\rangle), \end{aligned} \quad (2.7)$$



Here, the dashed line indicates how to consider $A_a^x \stackrel{\text{def}}{=} \langle \chi_x | A_a | \chi_x \rangle$, respectively $C_c^z \stackrel{\text{def}}{=} \langle \zeta_z | C_c | \zeta_z \rangle$, as operators acting on one part of the bipartite state $|\psi_{AB}\rangle$, respectively $|\psi_{BC}\rangle$. By $\sum_a A_a = \mathbb{1}$ and normalization of $|\chi_x\rangle$, it follows that $\sum_a A_a^x = \mathbb{1}$ for all x ; similarly, $\sum_z C_c^z = \mathbb{1}$ for all z . By definition, (2.4) can then be written as

$$p(a, b, c | x, z) = (\langle \psi_{AB} | \otimes \langle \psi_{BC} |) (A_a^x \otimes B_b \otimes C_c^z) (|\psi_{AB}\rangle \otimes |\psi_{BC}\rangle). \quad (2.8)$$

This is desired quantum representation of p in a BGP scenario.

Conversely, we start from a correlation $p(a, b, c, x, z)$ of the form (2.8). As sources between A and X and between C and Z, we again take hidden variables defined by $\lambda_{XA} = x$ and $\lambda_{CZ} = z$; again, the protocol of X and Z is simply to announce the values of these variables as their outcome. Only the sources between A and B and between B and C are taken to be quantum and produce, respectively, the bipartite states $|\psi_{AB}\rangle$ and $|\psi_{BC}\rangle$ of (2.8). The measurement protocol conducted by A is similar to above: measure λ_{XA} , use the result as the choice of setting for the subsequent measurement on $|\psi_{AB}\rangle$, and then announce both outcomes as the total outcome. This protocol can be interpreted as measuring a single POVM given by

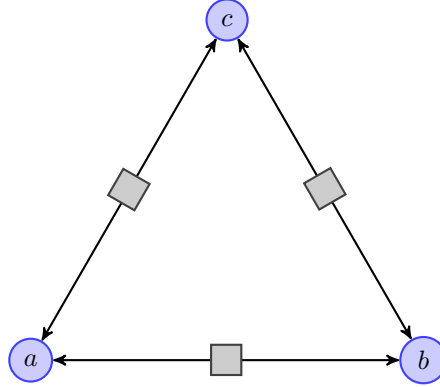
$$|x\rangle\langle x| \otimes A_a^x \mapsto (x, a),$$

where the left-hand side is a POVM element indexed by x and a , and the right-hand side denotes the resulting outcome announced by A. The analogous POVM is measured by C. By construction, this reproduces both the desired conditional distribution (2.8) and the marginal distribution $p(x, z) = p(x)p(z)$, and therefore also the whole distribution $p(x, a, b, c, z)$. \square

Due to this theorem, we can regard P_5 as the analogue of the BGP scenario within our formalism.

However, this is not yet the end of the story; our new point of view provides more than just a reformulation of familiar things. Let us imagine that party Z, in the P_5 scenario, has failed to collect data. Or that we disregard Z's measurement for some other reason. Then, we can regard the remaining parties X, A, B, C as forming a P_4 scenario and apply Theorem 2.4 to the distribution $p(x, a, b, c)$, with c now playing the role of y . In this way, the P_4 scenario is a natural *subscenario* of P_5 . This is an observation which does not make sense in the standard formalism.

The triangle scenario C_3 . Our next example, first proposed in [BRGP12, Sec.VI], is the correlation scenario illustrated in Figure 3. It consists of three parties of which each two share a common source. We will see in Corollary 3.10 that it is the smallest scenario in which non-classical

FIGURE 3. The correlation scenario C_3 .

correlations exist. In this subsection, we prove the existence of non-classical quantum correlations in C_3 .

We find this scenario especially appealing both due to its symmetry and due to its appearance in the study of inference of common ancestors [SA10]; see below. Since the main ideas concerning correlation scenarios should already have become clear in the last two examples, we now increase the pace a bit.

DEFINITION 2.10. *A correlation in C_3 is a distribution $p(a, b, c)$. (It is not required to satisfy any particular constraint.)*

This definition seems reasonable to us since, in general, one cannot expect any two of the variables (a, b, c) to be independent.

EXAMPLE 2.11. If all three variables take values in $\{0, 1\}$, then

$$p(a = b = c = 0) = \frac{1}{2}, \quad p(a = b = c = 1) = \frac{1}{2}$$

defines a correlation. We call this the *perfect correlation* since all three variables are random, but perfectly correlated.

DEFINITION 2.12. *A correlation $p(a, b, c)$ is classical in the C_3 scenario if and only if it can be written in the form*

$$p(a, b, c) = \int_{\lambda_{AB}, \lambda_{BC}, \lambda_{CA}} p(a|\lambda_{CA}, \lambda_{AB})p(b|\lambda_{AB}, \lambda_{BC})p(c|\lambda_{BC}, \lambda_{CA})p(\lambda_{AB})p(\lambda_{BC})p(\lambda_{CA}) \quad (2.9)$$

for appropriate (conditional) distributions $p(a|\lambda_{CA}, \lambda_{AB})$, $p(b|\lambda_{AB}, \lambda_{BC})$, $p(c|\lambda_{BC}, \lambda_{CA})$, $p(\lambda_{AB})$, $p(\lambda_{BC})$, $p(\lambda_{CA})$.

Classical correlations in C_3 are monogamous in the following sense:

PROPOSITION 2.13. *Let $p(a, b, c)$ be classical. If $p(a = c) = 1$, then a is independent of λ_{AB} .*

Intuitively, this is because in order to create these perfect correlations between a and c , the outcome a cannot depend on λ_{AB} . In particular, this implies that there cannot be any correlations between a and b . Rigorously, the proof technique is the same as the one used in the proof of this inequality relating Shannon entropy and mutual information, which can be regarded as a monogamy inequality:

LEMMA 2.14. *Let $p(a, b, c)$ be classical. Then*

$$I(a : b) + I(a : c) \leq H(a). \quad (2.10)$$

The interpretation of this is a kind of monogamy: a can share strong correlations with only b or c , but not with both. In particular, this inequality shows that the perfect correlation of Example 2.11 is not classical.

PROOF OF PROPOSITION 2.13 AND LEMMA 2.14. The present proof concerns the case that the hidden variables are discrete; see A.4 for the general case.

Since a and b are conditionally independent given λ_{AB} , and similarly for a and c , the data processing inequality can be used to bound the left-hand side of (2.10) by

$$I(a : b) + I(a : c) \leq I(a : \lambda_{AB}) + I(a : \lambda_{CA}) = 2H(a) + H(\lambda_{AB}) + H(\lambda_{CA}) - H(a\lambda_{AB}) - H(a\lambda_{CA}).$$

Submodularity of Shannon entropy guarantees that $H(a\lambda_{AB}) + H(a\lambda_{CA}) \geq H(a) + H(a\lambda_{AB}\lambda_{CA})$, which can be applied here to obtain

$$I(a : b) + I(a : c) \leq H(a) + H(\lambda_{AB}) + H(\lambda_{CA}) - H(a\lambda_{AB}\lambda_{CA}) \leq H(a) + I(\lambda_{AB} : \lambda_{CA}).$$

Since $I(\lambda_{AB} : \lambda_{CA}) = 0$, the claim of Lemma 2.14 follows.

Concerning Proposition 2.13, its assumption implies $I(a : c) = H(a)$; the sequence of inequalities derived in this proof then guarantees that $I(a : \lambda_{AB}) = 0$, as was to be shown. \square

COROLLARY 2.15. *Let $p(a, b, c)$ be classical and f, g functions such that $f(a)$ and $g(c)$ are defined. If $p(f(a) = g(c)) = 1$, then $f(a)$ and $g(c)$ are independent of λ_{AB} .*

PROOF. The assumptions imply that $p(f(a), b, g(c))$ is also a classical correlation in C_3 . Now the claim follows from Proposition 2.13. \square

THEOREM 2.16. *There exist non-classical quantum correlations in C_3 .*

PROOF. We take $|\psi\rangle$ to be a bipartite two-qubit state which violates the CHSH inequality [CHSH69] with respect to measurements in the two bases $\{|\phi_0\rangle, |\phi_1\rangle\}$, $\{|\omega_0\rangle, |\omega_1\rangle\}$, which are the same for both parties.

The quantum correlations we consider in C_3 are obtained as follows. We take A and B to share $|\psi\rangle$, while A and C as well as B and C share either a maximally entangled state

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} \quad (2.11)$$

or, equivalently, a classically correlated mixed state

$$\frac{1}{2} (|00\rangle\langle 00| + |11\rangle\langle 11|) \quad (2.12)$$

of two qubits. The purpose of these states is simple: it obsoletes free will in that A and B first measure the system they receive from the source shared with C in the $\{|0\rangle, |1\rangle\}$ -basis and use the resulting outcome as a measurement setting on $|\psi\rangle$; this is similar to how the proofs of Theorems 2.4 and 2.9 work. A and B announce the outcomes of both measurements as their total outcome. Similarly, we take C to apply the $\{|0\rangle, |1\rangle\}$ -measurement on each of his qubits, so that C knows the measurement “setting” used by A and B. He announces both of them as his outcome c . We regard the two bits announced by each party as the outcome of a single four-outcome measurement. The resulting correlation $p(a, b, c)$ is a probability distribution on 4^3 outcomes which does not depend on whether (2.11) or (2.12) is used.

More formally, we can define the measurements as follows: both **A** and **B** measure in the following basis and announce respective outcomes:

$$\begin{aligned} |0\rangle\langle 0| \otimes |\phi_0\rangle\langle \phi_0| &\mapsto (0, 0), & |0\rangle\langle 0| \otimes |\phi_1\rangle\langle \phi_1| &\mapsto (0, 1), \\ |1\rangle\langle 1| \otimes |\omega_0\rangle\langle \omega_0| &\mapsto (1, 0), & |1\rangle\langle 1| \otimes |\omega_1\rangle\langle \omega_1| &\mapsto (1, 1), \end{aligned}$$

while **C** simply measures both his qubits in the standard basis and announces both results.

It needs to be proven that these correlations are non-classical in C_3 . This is guaranteed by the monogamy property of Corollary 2.15: since **C** has perfect information about the “settings” employed by **A** and **B**, these “settings” are necessarily independent of λ_{AB} . This simulates the free will (“ λ -independence”) required for a standard Bell test to apply. The hidden variable λ_{AB} in any potential classical model would therefore have to function exactly like a hidden variable in a standard Bell scenario, which is guaranteed to be impossible due to the Bell inequality violation. \square

These arguments apply in the same way to a construction of a non-classical quantum correlations from a Bell inequality violation in any bipartite Bell scenario.

Although this class of examples proves the theorem, we do not find such examples satisfying since they are again based on a Bell test in the standard sense. It is difficult to regard them as entirely new kinds of non-classicality. Nevertheless, we find it surprising that non-classical quantum correlations exist in C_3 even in the case when *only one* of the sources produces entanglement. We had not expected this at all when we started thinking about the C_3 scenario.

PROBLEM 2.17. *Find an example of non-classical quantum correlations in C_3 together with a proof of its non-classicality which does not hinge on Bell’s Theorem.*

In order to find more examples of non-classical quantum correlations in C_3 , it would be helpful to have inequalities bounding the set of classical correlations and violated by some quantum correlations. Unfortunately, our proof of Theorem 2.16 does provide inequalities only conditional on the perfect correlations required between **A** and **C** and between **B** and **C**. However, we expect that our idea can be used to derive unconditional inequalities, if one knows bounds on the maximal classical value of a Bell inequality as a function of the correlation between the measurement settings and the hidden variable. We expect that such bounds can be derived by considerations similar to those of [BG11] and/or [CR12] or may even be implicitly contained in these works.

Before moving on to the next example of a correlation scenario, we return briefly to the work of Steudel and Ay [SA10] on the inference of common ancestors. So, what is a “common ancestor”?

If one makes certain (say, real-world macroscopic) observations a and b , repeats them many times in order to gather statistics, and detects a correlation between these, then one can conclude that a and b need to have a *common ancestor*: there needs to be some quantity or property λ such that both a and b depend on λ , and λ is not deterministic; this includes the possibilities $\lambda = a$ and $\lambda = b$ as degenerate cases. This λ is a common ancestor of a and b in the sense of a preexisting condition on which both a and b depend.

This is Reichenbach’s principle of common cause [RR56, Ebe08]; it is based on the premise that good models of the world adhere to assumption (III’) in the sense that a good model should predict a and b to be independent, unless there is some previously occurring event causally connected to both variables, i.e. a common ancestor.

Now what if one does the same for three observations a , b , c ? How can one conclude that there is a common ancestor λ on which all three of them depend? Or for any number $n \in \mathbb{N}$ of observations? Among other things, it has been shown in [SA10] that the entropy of the common

ancestors is lower bounded by a certain linear combination of the joint entropy and the single-variable entropies; therefore, strict positivity of that linear combination witnesses the necessity of a common ancestor. See also [Ay09] for related work providing a generalization and quantification of Reichenbach’s principle.

Let us consider the particular case of $n = 3$ variables. Then the main observation is that the causal structure of the C_3 scenario is precisely the null hypothesis: if no common ancestor exists, then there can at most be common ancestors for every pair of variables, but not for all three variables together. Therefore, *if no common ancestor exists, then $p(a, b, c)$ is a classical correlation in the C_3 scenario*. Figure 3 coincides with [SA10, Figure 1]. The results of Steudel and Ay for this particular case state that if $p(a, b, c)$ is classical, then

$$H(a) + H(b) + H(c) \leq 2H(abc), \quad (2.13)$$

where $H(abc)$ is the entropy of the joint distribution. Intuitively, if this inequality is violated, then the joint entropy is relatively small in comparison to the single-variable entropies, implying the existence of strong correlations between the variables and therefore of a common ancestor.

Writing out our inequality (2.10) in terms of joint entropies, one obtains

$$H(a) + H(b) + H(c) \leq H(ab) + H(ac),$$

which is an improvement over (2.13) since the right-hand side is bounded by $2H(abc)$. In particular, a violation

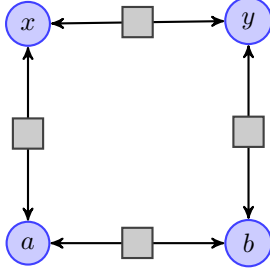
$$H(a) + H(b) + H(c) > H(ab) + H(ac) \quad (2.14)$$

successfully witnesses the necessity of a common ancestor in strictly more cases than (1.1).

In the case of $n > 3$ variables, it is still true that the null hypothesis of non-existence of a common cause corresponds to classicality in the appropriate correlation scenario: for the necessity of a common ancestor of some $(k + 1)$ -element subset of n variables, the null hypothesis is that at most each k -tuple has common ancestor(s). Roughly speaking, it is enough to consider only those ancestors which themselves do not have any parents: all the randomness creation can be delegated to those without changing the distribution of the observed variables, while all other nodes then carry out deterministic information processing; compare Remark 2.3 and A.3. Then each such initial node can be replaced by a source connecting to at most k observed variables, and the deterministic information processing can as well be delegated to the measurement nodes, again without changing the distribution of outcomes. Therefore, this corresponds to a classical correlation in the correlation scenario defined by n measurements in which each k -tuple of measurements is allowed to share a source. Conversely, it is clear that every such classical correlation represents a joint distribution of n variables which can be modelled without a common ancestor for any $(k + 1)$ -tuple. To summarize, the given joint distribution is a classical correlation in this scenario if and only if the joint distribution can be obtained from a Bayesian network in which no $(k + 1)$ -element subset of the given variables has a common ancestor.

However, at the moment we do not know how to generalize our inequality (2.14) to these cases, and refer once again to [SA10] for the current state of the art.

The square scenario C_4 . Another interesting correlation scenario is the *square scenario* illustrated in Figure 4. In this case, the underlying graph is C_4 , the cycle graph on four vertices. It can be regarded as P_4 (Figure 1) equipped with an additional source between X and Y . Along the lines of Theorem 2.4, this would suggest that correlations $p(a, b, x, y)$ in C_4 should be interpretable as arising from a Bell scenario together with correlations between the measurement settings. However, the forthcoming Proposition 2.20 will show that this intuition is false.

FIGURE 4. The correlation scenario C_4 .

DEFINITION 2.18. A correlation p in the C_4 scenario is a distribution $p(a, b, x, y)$ whose marginals factorize as

$$p(a, y) = p(a)p(y), \quad p(b, x) = p(b)p(x).$$

DEFINITION 2.19. A correlation $p(a, b, x, y)$ is classical in the C_4 scenario if and only if it can be written in the form

$$p(a, b, x, y) = \int_{\lambda_{AB}, \lambda_{BY}, \lambda_{YX}, \lambda_{XA}} p(a|\lambda_{XA}, \lambda_{AB})p(b|\lambda_{AB}, \lambda_{BY})p(y|\lambda_{BY}, \lambda_{YX})p(x|\lambda_{YX}, \lambda_{XA})p(\lambda_{AB})p(\lambda_{BY})p(\lambda_{YX})p(\lambda_{XA}) \quad (2.15)$$

for appropriate (conditional) distributions $p(a|\lambda_{XA}, \lambda_{AB})$, $p(b|\lambda_{AB}, \lambda_{BY})$, $p(y|\lambda_{BY}, \lambda_{YX})$, $p(x|\lambda_{YX}, \lambda_{XA})$, $p(\lambda_{AB})$, $p(\lambda_{BY})$, $p(\lambda_{YX})$, $p(\lambda_{XA})$.

PROPOSITION 2.20. There are classical correlations $p(a, b, x, y)$ in the C_4 scenario such that the associated conditional distribution $p(a, b|x, y)$ is signaling.

PROOF. We start from any classical correlation $p_0(x, a, b, y)$ in the P_4 scenario. In particular, by Theorem 2.4, $p_0(a, b|x, y)$ does not violate any Bell inequality. We now apply the relabeling

$$a \longleftrightarrow x, \quad b \longleftrightarrow y$$

and take the resulting correlation to be $p(a, b, x, y)$. By construction, the resulting correlation $p(a, b, x, y)$ is classical in C_4 . By construction, $p(x, y|a, b)$ does not violate a Bell inequality. The conditional distribution

$$p(a, b|x, y) = p(x, y|a, b) \cdot \frac{p(x, y)}{p(a)p(b)}$$

then is precisely the time reversal, in the sense of Coecke and Lal [CL12], of the classical no-signaling box $p(x, y|a, b)$ with respect to $p(a, b) = p(a)p(b)$ as its distribution of settings. It was shown in [CL12] that there exist $p_0(a, b|x, y)$ for which this time reversal is necessarily signaling. \square

In particular, Proposition 2.20 shows that the conditional distribution $p(a, b|x, y)$ associated to a classical correlation $p(a, b, x, y)$ in C_4 may violate Bell inequalities.

Any classical (resp. quantum) correlation in a bipartite Bell scenario can be turned into a classical (resp. quantum) correlation in the C_4 scenario in four different ways: one of the four edges of C_4 needs to be designated as the Bell scenario's source, while the source corresponding to the opposite edge does nothing at all.

THEOREM 2.21. There exist non-classical correlations in C_4 .

PROOF. We define the correlation $p(a, b, x, y)$ by taking $p(a, b|x, y)$ to be a Popescu-Rohrlich box [PR94] and $p(x, y)$ to be the uniform distribution. More concretely, all four outcomes are bits $a, b, x, y \in \{0, 1\}$ with the table of joint probabilities given by:

| | | | | | | |
|------------|--------|---------------|---------------|---------------|---------------|--------|
| | | $(x, a) =$ | | | | |
| | | (0, 0) | (0, 1) | (1, 0) | (1, 1) | |
| $(y, b) =$ | (0, 0) | $\frac{1}{8}$ | 0 | $\frac{1}{8}$ | 0 | (2.16) |
| | (0, 1) | 0 | $\frac{1}{8}$ | 0 | $\frac{1}{8}$ | |
| | (1, 0) | $\frac{1}{8}$ | 0 | 0 | $\frac{1}{8}$ | |
| | (1, 1) | 0 | $\frac{1}{8}$ | $\frac{1}{8}$ | 0 | |

We now use a Hardy-type [Har93] argument in order to show that this correlation is not classical. For the sake of contradiction, let us assume $p(a, b, x, y)$ to be classical with hidden variable distributions $p(\lambda_{AB}), p(\lambda_{BY}), p(\lambda_{YX}), p(\lambda_{XA})$; thanks to Remark 2.3, we can take the four outcomes to be deterministic functions of the hidden variables. We start by considering the case of discrete hidden variables. Then, there has to be a hidden variable combination

$$(\lambda_{AB}, \lambda_{BY}, \lambda_{YX}, \lambda_{XA}) = (\ell_{AB}, \ell_{BY}, \ell_{YX}, \ell_{XA})$$

occurring with positive probability, which produces the outcome $(a, b, x, y) = (0, 0, 0, 0)$; similarly, there has to be a hidden variable combination

$$(\lambda_{AB}, \lambda_{BY}, \lambda_{YX}, \lambda_{XA}) = (\kappa_{AB}, \kappa_{BY}, \kappa_{YX}, \kappa_{XA}),$$

occurring with positive probability, which produces the outcome $(a, b, x, y) = (1, 0, 1, 1)$. Then, the independence of sources guarantees that the hidden variable combination

$$(\lambda_{AB}, \lambda_{BY}, \lambda_{YX}, \lambda_{XA}) = (\kappa_{AB}, \kappa_{BY}, \ell_{YX}, \kappa_{XA}),$$

also has positive probability. Because of locality and determinism, it necessarily produces an outcome $(1, 0, \hat{x}, \hat{y})$; by (2.16), $\hat{x} = \hat{y} = 1$. Likewise, the hidden variable combination $(\ell_{AB}, \ell_{BY}, \ell_{YX}, \kappa_{XA})$ has positive probability, and produces some outcome of the form $(a', 0, 1, 0)$. Thanks to the form of (2.16), necessarily $a' = 0$. Similarly, the hidden variable combination $(\ell_{AB}, \kappa_{BY}, \ell_{YX}, \ell_{XA})$ must give the outcome $(0, 0, 0, 1)$. However, the hidden variable combination $(\ell_{AB}, \kappa_{BY}, \ell_{YX}, \kappa_{XA})$ then gives the outcome $(0, 0, 1, 1)$ with positive probability, a contradiction with (2.16).

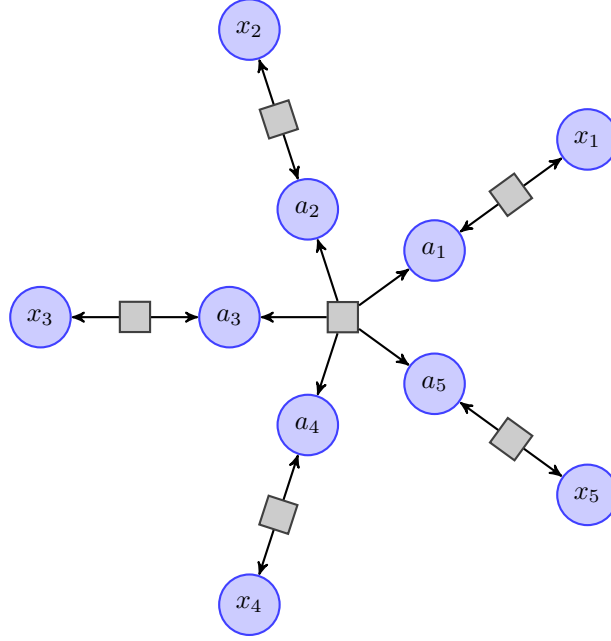
In the case of general (non-discrete) hidden variables, the same proof idea can be used, although the technical details are quite involved; see A.5. \square

PROBLEM 2.22. (1) Are there non-classical quantum correlations in C_4 ?

(2) Is there a simple way to characterize the classical correlations in C_4 ?

Scenarios with multipartite sources. So far, we have only considered example scenarios in which each source produces a pair of systems which it distributes among two parties. However, it is quite common to consider Bell scenarios involving a source that distributes systems among several parties [GHZ90]. The same can be easily done in our framework; an example scenario of this type is illustrated in Figure 5.

More generally, we want to consider the family of *multiarm scenarios* A_k indexed by the number of arms $k \in \mathbb{N}$; each arm consists of two parties sharing a bipartite source, and there is one k -partite source shared by all the parties obtained by choosing one party in each arm. Figure 5 represents the case $k = 5$, while $k = 2$ is the P_4 scenario of Figure 1.

FIGURE 5. The correlation scenario A_5 .

The following considerations are immediate generalizations of those of the P_4 scenario. Just as P_4 corresponds to a bipartite Bell scenario, A_k corresponds to a k -partite Bell scenario. We use “hat” notation like $a_1, \dots, \hat{a}_i, \dots, a_k$ as short for $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_k$.

DEFINITION 2.23. A correlation in the A_k scenario is a probability distribution $p(a_1, \dots, a_k, x_1, \dots, x_k)$ whose marginals factorize as

$$p(a_1, \dots, \hat{a}_i, \dots, a_k, x_1, \dots, x_k) = p(x_i) p(a_1, \dots, \hat{a}_i, \dots, a_k, x_1, \dots, \hat{x}_i, \dots, x_k). \quad \forall i \quad (2.17)$$

Repeated application of (2.17) implies $p(x_1, \dots, x_n) = p(x_1) \cdots p(x_n)$. Upon using this, and considering only those values x_i for which $p(x_i) > 0$, the condition (2.17) becomes equivalent to the equations

$$p(a_1, \dots, \hat{a}_i, \dots, a_n | x_1, \dots, x_n) = p(a_1, \dots, \hat{a}_i, \dots, a_n | x_1, \dots, \hat{x}_i, \dots, x_n),$$

which are formally identical to the no-signaling equations in a k -partite Bell scenario.

THEOREM 2.24. (1) A correlation p is classical in A_k if and only if the associated conditional distribution $p(a_1, \dots, a_k | x_1, \dots, x_k)$ is classical in the Bell scenario sense.
 (2) A correlation $p(a, b, x, y)$ is quantum in A_k if and only if the associated conditional distribution $p(a_1, \dots, a_k | x_1, \dots, x_k)$ is quantum in the Bell scenario sense

PROOF. Analogous to the proof of Theorem 2.4. □

3. General theory of correlation scenarios

We now adopt a more abstract point of view. Looking at the previous examples, one should come to the conclusion that a general definition of correlation scenario should define the data of a

correlation scenario to consist of a set of measurements (= parties = observers) M , a set of sources S , and a relation $C \subseteq S \times M$ between sources and measurements, where we write $(s, m) \in C$ also as sCm and read it as “ s connects to m ”. As before, the physical picture is that each source sends out one physical system to each party it connects to, and each party conducts a fixed measurement on the collection of systems it receives from the sources it is connected to. The temporal (or rather causal) structure of such a scenario consists of a primary layer of sources and a secondary layer of measurements. In [FS12], we will go beyond this “two-layer” approach and consider a vastly more general formalism allowing for any kind of causal structure.

Finally, a correlation scenario should also specify how many possible outcomes each measurement has. For simplicity, we take this to be the same number $d \in \mathbb{N}$ for all measurements. We usually omit mention of d and regard it as implicitly defined through the correlation: given the joint outcome distribution, d can be taken to be equal to the highest number of actually occurring outcomes over all measurements.

DEFINITION 3.1. *A correlation scenario is a quadruple (S, M, C, d) consisting of a finite set of sources S , a finite set of measurements M , a relation $C \subseteq S \times M$ (read: “connects”) and a natural number $d \in \mathbb{N}$. The relation is required to satisfy the conditions*

- (1) $(s_1 C m \Rightarrow s_2 C m \forall m) \iff s_1 = s_2$
- (2) $(s C m_1 \Leftarrow s C m_2 \forall s) \iff m_1 = m_2$

These two conditions are to be interpreted as follows: if source s_2 connects to each measurement to which also s_1 connects, then s_1 is redundant. Therefore, we may assume without loss of generality that such redundancies do not occur: if s_1 connects to a subset of the measurements to which s_2 connects, or to exactly the same measurements, then $s_1 = s_2$. Similarly, if there are two measurements which connect to exactly the same set of sources, then we may replace both measurements by a single one. Therefore, we assume without loss of generality that if m_1 and m_2 connect to the same set of sources, then $m_1 = m_2$.

The scenarios depicted in Figures 1–5 are exactly of this form: the circles represent M , the boxes form S , and the arrows define C .

DEFINITION 3.2. *A hypergraph $G = (V, E)$ consists of a finite set of vertices V and a set of edges $E \subseteq 2^V$, i.e. every edge $e \in E$ is a subset $e \subseteq V$.*

The combinatorial data of Definition 3.1 can equivalently be specified in terms of a hypergraph. One obtains a hypergraph from a correlation scenario (S, M, C, d) by using the vertex set $V = M$ and introducing one edge for each source which contains exactly those vertices (= measurements) to which the source connects. Formally, the resulting set of edges is

$$E = \{\{r \in P : sCr\}, s \in S\}.$$

Then the two requirements of Definition 3.1 translates into the properties

- (1) G is an anti-chain: there is no edge which is contained in a different one.
- (2) There are no two different vertices which belong to exactly the same set of edges.

Conversely, every hypergraph with these properties defines a correlation scenario in the obvious way: vertices become measurements, and every edge defines a source which connects to all those measurements contained in the edge.

For now, we stick with this hypergraph picture. In other words, we identify a source with the set of measurements that it connects to. For the following, we fix a hypergraph $G = (V, E)$, satisfying (1), (2), together with some $d \in \mathbb{N}$ for the number of possible outcomes. We take this

data to represent any correlation scenario. We write $V = \{v_1, \dots, v_n\}$ and associate to each vertex v_i a random variable, representing the measurement outcome distribution, which we also denote by v_i . The following definition generalizes the Definitions 2.1, 2.7, 2.10, 2.18 and 2.23.

DEFINITION 3.3. *A correlation p in G is a probability distribution $p(v_1, \dots, v_n)$ such that for every pair of subsets $U, W \subseteq V$ which are not connected in G (i.e. $\nexists e \in E$ with $U \cap e \neq \emptyset \wedge W \cap e \neq \emptyset$),*

$$p(u_1, \dots, u_{|U|}, w_1, \dots, w_{|W|}) = p(u_1, \dots, u_{|U|})p(w_1, \dots, w_{|W|}).$$

where $U = \{u_1, \dots, u_{|U|}\}$ and $W = \{w_1, \dots, w_{|W|}\}$.

It follows immediately that the same property not only holds for a pair of subsets of V , but for any number of pairwise not connected subsets.

PROBLEM 3.4. *For every standard Bell scenario, there is a general probabilistic theory [Bar07] which reproduces all no-signaling correlations in that scenario². Is this also true that for every correlation scenario? If not, are there other frameworks beyond general probabilistic theories in which this would be the case? Or would that mean that our Definition 3.3 is too lax?*

In a hidden variable model, each source $e \in E$ is described by a hidden variable λ_e with some distribution $p(\lambda_e)$. The locality assumption (II) then allows an outcome v_i to depend on all the sources connected to v_i ; we write $\Lambda_i = \{\lambda_e; v_i \in e\}$ for the set of hidden variables associated to all those sources.

DEFINITION 3.5. *A correlation p in G is classical if there are distributions $p(\lambda_e)$ and conditional distributions $p(v_i|\Lambda_i)$ such that*

$$p(v_1, \dots, v_n) = \int_{\{\lambda_e, e \in E\}} \prod_{v_i \in V} p(v_i|\Lambda_i) \prod_{e \in E} p(\lambda_e) \quad (3.1)$$

See 3.5 for the precise measure-theoretical definition. It is a simple exercise to check that every classical correlation is indeed a correlation as in Definition 3.3.

PROBLEM 3.6. *Under which conditions on G are all correlations classical?*

A class of scenarios in which all correlations are trivially classical is this:

PROPOSITION 3.7. *If there is a source in G connecting to all vertices, i.e. if $E = \{V\}$, then every distribution $p(v_1, \dots, v_n)$ is a classical correlation in G .*

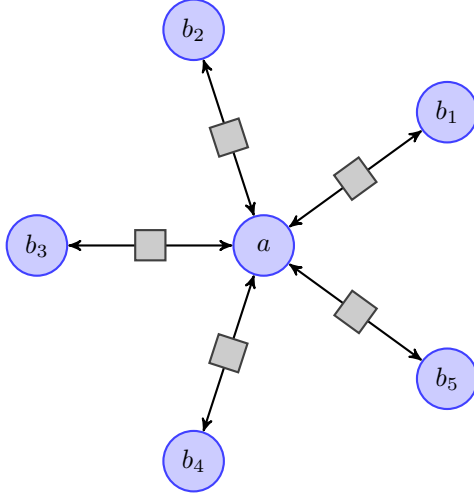
PROOF. The hidden variable carried by the common source can be taken to be $\lambda = (v_1, \dots, v_n)$ itself: in each run of the experiment, it selects a joint outcome (v_1, \dots, v_n) according to the desired distribution, sends this joint outcome as a hidden variable λ to all measurements. The outcome v_i is then defined to be the i th component of λ . \square

Using our previous analysis of example scenarios together with a bit of graph theory, we can answer Problem 3.6 at least in the case of bipartite sources, i.e. when the hypergraph $G = (V, E)$ is a (undirected, simple) graph. The relevant class of correlation scenarios turns out to be the class of *star scenarios* S_k indexed by the number $k \in \mathbb{N}$. The star graph S_k is defined to have vertices $V = \{a, b_1, \dots, b_k\}$ and one edge between a and every b_i , i.e.

$$E = \{\{a, b_1\}, \dots, \{a, b_k\}\}.$$

See Figure 6.

²For example, take the corresponding no-signaling polytope as the state space of the total system.

FIGURE 6. The correlation scenario S_5 .

THEOREM 3.8. *If the hypergraph G is a graph, then all correlations in G are classical if and only if G is a star graph or a disjoint union of star graphs.*

We begin the proof with a lemma.

LEMMA 3.9. *Let $G = (V, E)$ be a connected simple graph. If G is not a star graph, then G has some induced subgraph which is a C_3 , C_4 or P_4 .*

PROOF. We use induction on $n = |V|$. For $n \leq 3$, the statement is clear, since the only connected graphs at most three vertices are C_3 and the star graphs $P_1 = S_0$, $P_2 = S_1$ and $P_3 = S_2$. For $n \geq 4$, we start with G and assume that G does not contain any induced C_3 , C_4 or P_4 . We now select any induced subgraph on $n - 1$ vertices. By the induction assumption, this subgraph is a star graph with some central vertex $a \in V$ and leaves $b_1, \dots, b_{n-2} \in V$. For the induction step, we ask: how can the additional vertex $c \in V$ be connected to a, b_1, \dots, b_{n-2} ? An edge from c to a together with one from c to some b_i would give rise to an induced subgraph of type C_3 ; no edge to a but an edge to some b_i would give rise to an induced subgraph of type C_4 or P_4 . Therefore, c cannot share an edge with any b_i . Then due to connectedness, it needs to share an edge with a , which turns it into another leaf of the star. \square

PROOF OF THEOREM 3.8. If G is not a star graph or a disjoint union of star graphs, then the lemma guarantees that G contains an induced C_3 , C_4 or P_4 . Any correlation on such an induced subgraph can be extended to a correlation on G by taking the measurements associated to the additional vertices to have a deterministic outcome. Any hidden variable model of this extension can be restricted to a hidden variable model of the original correlation on the subgraph; in other words, if the original correlation is non-classical, then so is the extension. The existence of non-classical correlations on G now follows from Theorems 2.4, 2.16 and 2.21.

We now consider the case that $G = (V, E)$ is a star graph. This means that $V = \{a, b_1, \dots, b_n\}$, where a is the central vertex sharing an edge with each b_i , and there are no other edges. It follows

from Definition 3.3 that a correlation on G is a distribution $p(a, b_1, \dots, b_n)$ satisfying

$$p(a, b_1, \dots, b_n) = p(a|b_1, \dots, b_n) \prod_{i=1}^n p(b_i).$$

Defining hidden variables as $\lambda_{AB_i} = b_i$ shows that p is indeed classical. \square

Since C_3 is the only hypergraph on 3 vertices for which not all measurements share a common source and which is not a star graph, we obtain as a direct consequence:

COROLLARY 3.10. *C_3 is the smallest scenario in which non-classical correlations exist.*

If Problem 2.22(1) has a positive answer, then “non-classical” can also be replaced in Theorem 3.8 and Corollary 3.10 by “non-classical quantum”.

For Bell scenarios, it is an open problem whether all quantum correlations in a fixed Bell scenario can be achieved quantum-mechanically in terms of quantum states on a Hilbert space of fixed dimension. Numerical evidence suggests that this is not the case in general [PV10]. Due to Theorem 2.4, this question as well as the numerical evidence automatically transfer to the P_4 scenario. The analogous question for the classical case is: how many values for the hidden variable(s) are required in order to simulate all classical correlations? In a Bell scenario, this is easily seen to be a finite number since the set of classical correlations is a convex polytope with the deterministic correlations as extremal points, so that Carathéodory's Theorem gives an explicit bound on the number of hidden variable values needed. However, in our more general formalism, the answer to the same question is not at all clear.

PROBLEM 3.11. *Are there correlation scenarios in which no finite number of values for the hidden variables is enough for obtaining all classical correlations with a given number of outcomes?*

Due to Theorem 2.4, we know that a finite number is sufficient in the case of P_4 . The natural next step will be to consider this problem for C_3 , where it already seems difficult.

PROBLEM 3.12. *Can the set of classical correlations be described by a finite number of polynomial inequalities?*

This is in fact related to Problem 3.11:

PROPOSITION 3.13. *Let G be a correlation scenario with a fixed number of outcomes for each measurement. If a finite number of hidden variable value suffices in G to obtain all classical correlations, then the set of classical correlations in G can be described in terms of a finite number of polynomial inequalities.*

PROOF. If $k \in \mathbb{N}$ hidden variable values are enough to simulate all classical correlations, then a distribution over these values is specified by $k - 1$ real numbers satisfying k linear inequalities. Similarly, a conditional distribution $p(v_i|\Lambda_i)$ is specified by a certain finite number of real variables satisfying certain linear inequalities. The question of whether a given correlation is classical is then equivalent to asking whether these real variables can be chosen in such a way that they satisfy these linear inequalities and reproduce the given p via (3.1). In other words, it boils down to deciding whether a given system of polynomial inequalities, containing the $p(v_1, \dots, v_n)$ as parameters, has a solution over \mathbb{R} .

Thanks to Tarski's real quantifier elimination [Tar51], this system of polynomial inequalities is solvable if and only if $p(v_1, \dots, v_n)$ itself satisfies certain polynomial inequalities which can in principle be computed explicitly. \square

Besides the trivial case of star graph scenarios, the only cases for which we know a positive answer to Problem 3.11, and therefore Problem 3.12, are the P_4 scenario (Theorem 2.4 and [Fin82]) and the P_5 scenario (Theorem 2.9 and [BGP10]).

We have already noted in Remark 2.6 that the set of classical correlations is not convex in general. So one may wonder:

PROBLEM 3.14. *What is the shape of the set of classical correlations? Can it have a non-trivial topology, or is it always homeomorphic to a ball of the appropriate dimension? If yes, what is this dimension? If no, is the set nevertheless contractible, or can it have “holes”? Is it always simply connected? What about the analogous questions for the set of quantum correlations?*

At the moment, we can only offer a very simple observation concerning these topological questions:

PROPOSITION 3.15. *Let $G = (V, E)$ be any correlation scenario with the number of outcomes of each measurement fixed to some $d \in \mathbb{N}$. Then the set of classical correlations is path-connected.*

PROOF. Given classical correlations $p_0(v_1, \dots, v_n)$ and $p_1(v_1, \dots, v_n)$ on G , we describe how to construct an explicit 1-parameter family of correlations continuously interpolating between these two. The assumption of classicality means that there are hidden variable distributions $p(\lambda_1^0), \dots, p(\lambda_m^0)$ and $p(\lambda_1^1), \dots, p(\lambda_m^1)$ together with the appropriate conditional distributions $p(v_i|\Lambda_i^0)$ and $p(v_i|\Lambda_i^1)$ such that

$$p_0(v_1, \dots, v_n) = \int_{\{\lambda_e; e \in E\}} \prod_{v_i \in V} p(v_i|\Lambda_i^0) \prod_{e \in E} p(\lambda_e^0),$$

$$p_1(v_1, \dots, v_n) = \int_{\{\lambda_e; e \in E\}} \prod_{v_i \in V} p(v_i|\Lambda_i^1) \prod_{e \in E} p(\lambda_e^1),$$

We now define a continuous family of classical correlations indexed by a parameter $t \in [0, 1]$. These as hidden variables the pairs $\lambda_e = (\lambda_e^0, \lambda_e^1)$ with distribution $p(\lambda_e) = p(\lambda_e^0, \lambda_e^1) \stackrel{\text{def}}{=} p(\lambda_e^0)p(\lambda_e^1)$.

For every $t \in [0, 1]$, we define a new conditional distribution for each random variable v_i ,

$$p_t(v_i|\Lambda_i) = (1-t) \cdot p(v_i|\Lambda_i^0) + t \cdot p(v_i|\Lambda_i^1), \quad (3.2)$$

and consider the resulting joint distribution

$$p_t(v_1, \dots, v_n) = \int_{\{\lambda_e; e \in E\}} \prod_{i \in V} p_t(v_i|\Lambda_i) \prod_{e \in E} p(\lambda_e). \quad (3.3)$$

By construction, this is a family of classical correlations depending continuously on t . For $t = 0$, the conditional distributions (3.2) do not depend on the λ_e^1 component of $\lambda_e = (\lambda_e^0, \lambda_e^1)$, so that the integration over λ_e^1 in (3.3) is trivial and the original $p_0(v_1, \dots, v_n)$ is reproduced. Similarly for $t = 1$. Then by continuity in t , the family p_t defines a continuous path of classical correlations between the two given classical correlations p_0 and p_1 . \square

For a similar proof idea, see [BRGP12, App. A.1].

We now return to the original picture of Definition 3.1 and consider some generalities on quantum correlations, starting with the rigorous definition. For a Hilbert space \mathcal{H} , we write $S(\mathcal{H})$ for the set of states on \mathcal{H} , i.e. positive trace-class operators of unit trace norm.

DEFINITION 3.16. *Let $G = (S, M, C, d)$ be a correlation scenario. A correlation $p(v_1, \dots, v_n)$ in G is quantum if the following data exist:*

- (1) for every connection $(s, m) \in C$, a Hilbert space $\mathcal{H}_{(s,m)}$;
- (2) for every source $s \in S$, a quantum state $\rho_s \in \mathfrak{S} \left(\bigotimes_{\{m : sCm\}} \mathcal{H}_{(s,m)} \right)$;
- (3) for every measurement $m \in M$, a POVM $\{\mathcal{F}_m\}$ with elements $\mathcal{F}_m \in \mathcal{B} \left(\bigotimes_{\{s : sCm\}} \mathcal{H}_{(s,m)} \right)$;

such that

$$p(m_1, \dots, m_n) = \text{tr} \left[\left(\bigotimes_{s \in S} \rho_s \right) \left(\bigotimes_{v_i \in V} \mathcal{F}_{v_i} \right) \right] \quad (3.4)$$

In this equation, both the left as well as the right tensor product evaluate to operators on $\bigotimes_{(s,m) \in C} \mathcal{H}_{(s,m)}$, but the two tensor products are taken with respect to different orders on C . We take it is as understood that these tensor products are taken to be reordered in such a way that the corresponding factors match.

We leave it to the reader to show that every classical correlation is also quantum.

PROPOSITION 3.17. *If all sources in a correlation scenario emit separable quantum states, then the resulting correlation is classical.*

PROOF. Here, we assume all Hilbert spaces to be finite-dimensional; see A.6 for the general case.

Carathéodory's Theorem guarantees the existence of some number $k \in \mathbb{N}$ such that every ρ_s can be decomposed as

$$\rho_s = \sum_{j=1}^k \mu_{s,j} \bigotimes_{\{m : sCm\}} \rho_{(s,m,j)} \quad (3.5)$$

for certain coefficients $\mu_{s,j} \geq 0$ with $\sum_{j=1}^k \mu_{s,j} = 1$ and certain states $\rho_{(s,m,j)} \in \mathfrak{S}(\mathcal{H}_{(s,m)})$. For each source s , we define its hidden variable λ_s to take values $j_s \in \{1, \dots, k\}$ with distribution $p(\lambda_s = j_s) = \mu_{s,j}$ and

$$p(m | \lambda_s = j_s \text{ for all } s \text{ with } sCm) \stackrel{\text{def}}{=} \text{tr} \left[\bigotimes_{s : sCm} \rho_{(s,m,j_s)} \mathcal{F}_m \right]. \quad (3.6)$$

This reproduces the correlation (A.1) for the states (3.5). Instead of verifying this formally, we would like to mention its interpretation as a concrete physical protocol. According to the decomposition (3.5), each source s can produce its state ρ_s by randomly generating λ_s , distribution according to the weights $\mu_{s,j}$, and preparing and sending the corresponding state $\rho_{(s,m,j)}$ to each party m for which sCm . In order to turn this into a completely classical protocol, we may shift the preparation of the states $\rho_{(s,m,j)}$ from the sources to the parties: if each party m knows the values of the hidden variables λ_s for all s with sCm , then this party itself can prepare the required states $\rho_{(s,m,j)}$ locally and measure them. In this way, only classical information λ_s has to be sent from the sources to the parties, and the parties' preparation and measurement can be considered as a single classical measurement on the λ_s 's given by the conditional probabilities (3.6). \square

PROBLEM 3.18. *Does every entangled quantum state display non-classical quantum correlations? I.e. can one obtain non-classical quantum correlations by choosing an appropriate correlation scenario and putting one copy of the state in each source? Does it help if each source also emits classical shared randomness in addition to the entangled state?*

Appendix A. Measure-theoretical technicalities and other nuisances

In the main text, we have assumed all our hidden variables to be discrete for the sake of readability. We drop this assumption here and consider the most general case: hidden variables can be arbitrary probability spaces. The following subsection are all referenced from the main text, so this appendix should be referred to only as needed.

A.1. What is a hidden variable? The literature knows examples of discrete hidden variables and continuous hidden variables. In standard Bell scenarios, Carathéodory's Theorem guarantees that considering discrete hidden variables is enough; unfortunately, we do not know whether this also holds for our case (Problem 3.11). Therefore, we should allow hidden variables which are as general as possible and require a definition which not only comprises discrete and continuous hidden variables, but also allows intermediate possibilities and even hidden variable with *more than* continuously many values.

Since the only successful general theory of (classical) randomness is the one based on the Kolmogorov axioms for probability measures and probability spaces, this is what seems to us to be the only reasonable general definition of hidden variable:

DEFINITION. A hidden variable is a probability space (Ω, \mathcal{E}, P) .

We think of the actual value of the hidden variable to be a random element $\lambda \in \Omega$ with distribution P . This is the most general kind of classical hidden variable we can imagine. It comprises both discrete and continuous variables as special cases as well as everything else, for example hidden variables with so many values that Ω has cardinality greater than the continuum.

A.2. Distributions conditional on hidden variables. Definitions 2.2, 2.8, 2.12, 2.19, 3.5 talk about outcome distributions conditional on one or several hidden variables. What does a conditional distribution, like $p(a|\lambda)$, mean when λ is not discrete?

There are several equivalent ways to answer this question. We have chosen the following one which is convenient in that it is partly formulated in terms familiar from quantum theory.

DEFINITION. Let $L^\infty(\Omega, \mathcal{E}, P)$ be the von Neumann algebra associated to (Ω, \mathcal{E}, P) . A distribution of a conditional on $\lambda \in \Omega$ is an assignment of some positive operator $\mathcal{O}_a^* = \mathcal{O}_a \in L^\infty(\Omega, \mathcal{E}, P)$, $\mathcal{O}_a \geq 0$, to every a such that $\sum_a \mathcal{O}_a = \mathbb{1}$.

The attentive reader will have noticed that this is nothing but a POVM in $L^\infty(\Omega, \mathcal{E}, P)$ indexed by a . Roughly speaking, each \mathcal{O}_a is a real-valued function on Ω whose values $\mathcal{O}_a(\lambda)$ represent the conditional probabilities $p(a|\lambda)$. For finite Ω with $\mathcal{E} = 2^\Omega$ and $P(\lambda) > 0$ for every $\lambda \in \Omega$, this intuition is exact; in general though, it has to be kept in mind that \mathcal{O}_a is not a single function, but rather a whole equivalence class of functions, such that expressions like

$$p(a) = \int_{\lambda} \mathcal{O}_a(\lambda) P(\lambda)$$

are well-defined in the sense that the value of the integral is independent of the choice of representative.

In general, a measurement a will depend on several hidden variables given by probability spaces $(\Omega_1, \mathcal{E}_1, P_1), \dots, (\Omega_n, \mathcal{E}_n, P_n)$. In this case, \mathcal{O} should be a POVM in the von Neumann algebra of the product probability space $(\prod_i \Omega_i, \prod_i \mathcal{E}_i, \prod_i P_i)$.

We now state Definition 3.5 again in the present language.

DEFINITION. Let $G = (V, E)$ be a correlation scenario. A correlation $p(v_1, \dots, v_n)$ in G is classical if the following data exist:

- (1) for every $e \in E$, a hidden variable λ_e given in terms of a probability space $(\Omega_e, \mathcal{E}_e, P_e)$;
- (2) Conditional probabilities $\mathcal{O}_a \in L^\infty(\Omega_{\Lambda_i}, \mathcal{E}_{\Lambda_i}, P_{\Lambda_i})$ where $\Lambda_i = \{\lambda_e; v_i \in e\}$ is the collection of hidden variables associated to all the sources connected to v_i , and $(\Omega_{\Lambda_i}, \mathcal{E}_{\Lambda_i}, P_{\Lambda_i})$ is the corresponding product probability space;

such that

$$p(v_1, \dots, v_n) = \int_{\{\lambda_e; e \in E\}} \prod_{v_i \in V} \mathcal{O}_{v_i}(\Lambda_i) \prod_{e \in E} dP(\lambda_e) \quad (\text{A.1})$$

In particular, this clarifies also the definitions of classical correlation in our example scenarios, Definitions 2.2 2.8, 2.12, 2.19

A.3. Hidden variables can be assumed deterministic. We have outlined in Remark 2.3 why the conditional distributions \mathcal{O}_a as used above can in fact taken to be deterministic. In our present picture, determinism means $\mathcal{O}_a^2 = \mathcal{O}_a$, i.e. that \mathcal{O}_a is a projection. This is equivalent to $\mathcal{O}_a(\lambda) \in \{0, 1\}$ for almost all $\lambda \in \Omega$ which corresponds to determinism in the form $p(a|\lambda) \in \{0, 1\}$.

We now turn the intuitive argument of Remark 2.3 into a rigorous proof sketch.

PROPOSITION. Let $G = (V, E)$ be a correlation scenario. If p is classical, then there exists a classical model for p in which all \mathcal{O}_{v_i} are projections.

PROOF. We show how to replace the \mathcal{O}_w 's by a projection for some fixed $w \in V$; the claim then follows from applying this procedure to every vertex $w \in V$. We start by choosing a source $e \in E$ which connects to v_1 and replace the given probability space $(\Omega_e, \mathcal{E}_e, P_e)$ by $\Omega'_e \stackrel{\text{def}}{=} \Omega_e \times [0, 1]$, which we take to be equipped with the product σ -algebra \mathcal{E}'_e and the product measure P'_e , where $[0, 1]$ carries the Lebesgue σ -algebra and measure; the second factor in this product represents the additional random number mentioned in Remark 2.3. We enumerate the possible outcomes as $w \in \{1, \dots, d\}$ for some $d \in \mathbb{N}$, and define

$$\mathcal{O}'_w : \Omega_e \times [0, 1] \rightarrow \{0, 1\}, \quad (\lambda, x) \mapsto \begin{cases} 1 & \text{if } \sum_{w'=1}^{w-1} \mathcal{O}_{w'}(\lambda) \leq x < \sum_{w'=1}^w \mathcal{O}_{w'}(\lambda) \\ 0 & \text{otherwise} \end{cases}$$

which is easily seen to represent a projection in $L^\infty(\Omega'_e, \mathcal{E}'_e, P'_e)$. The requirement $\sum_{w=1}^d \mathcal{O}'_w = \mathbb{1}$ holds by construction in $L^\infty(\Omega'_e, \mathcal{E}'_e, P'_e)$, i.e. up to a set of measure zero.

All \mathcal{O}_{v_i} with $v_i \neq w$ connecting to e we take to operate as before in the sense that we replace them by $\mathcal{O}'_{v_i}(\lambda, x) = \mathcal{O}_{v_i}(\lambda)$; all other sources $\neq e$ and all measurements not connected to e remain completely unchanged.

We leave it to the reader to verify that these replacements preserve the correlation. \square

A.4. General proof of Lemma 2.14. We follow essentially the same lines as in the discrete-variable proof of the main text. Since we do not know of a formulation of the data processing inequality for (relative) Shannon entropy on arbitrary probability spaces, and similarly for submodularity of entropy, we make our own definitions and derive our inequalities in analogy with the discrete case. We start with the first argument involving the data processing inequality. In order to obtain finite quantities, we need to work with conditional entropies, in which the hidden variables appear only as conditioning variables. For the sake of illustration, we start with the discrete-variable case, in which

$$H(a|\lambda_{AB}) = \sum_{a, \lambda_{AB}} f(p(a|\lambda_{AB})) p(\lambda_{AB}),$$

where we abbreviated $f(x) = -x \cdot \log x$, with $f(0) \stackrel{\text{def}}{=} 0$ as usual. Thanks to the conditional independence $p(a|\lambda_{\text{AB}}) = p(a|\lambda_{\text{AB}}, b)$ and concavity of f ,

$$\begin{aligned} H(a|\lambda_{\text{AB}}) &= \sum_{a,b,\lambda_{\text{AB}}} f(p(a|\lambda_{\text{AB}})) p(\lambda_{\text{AB}}) p(b|\lambda_{\text{AB}}) = \sum_{a,b} \sum_{\lambda_{\text{AB}}} f(p(a|\lambda_{\text{AB}}, b)) p(\lambda_{\text{AB}}|b) p(b) \\ &\leq \sum_{a,b} f\left(\sum_{\lambda_{\text{AB}}} p(a|\lambda_{\text{AB}}, b) p(\lambda_{\text{AB}}|b)\right) p(b) = \sum_{a,b} f(p(a|b)) p(b) = H(a|b). \end{aligned}$$

We now emulate this estimate in the general case by defining

$$H(a|\lambda_{\text{AB}}) \stackrel{\text{def}}{=} \sum_a \int_{\lambda_{\text{AB}}} f(\mathcal{O}_a(\lambda_{\text{AB}})) dP_{\text{AB}}(\lambda_{\text{AB}})$$

and noting that this is well-defined, thanks to $\mathcal{O}_a(\lambda_{\text{AB}}) \in [0, 1]$ a.s., and coincides with the standard definition in the discrete case. We rewrite this as

$$H(a|\lambda_{\text{AB}}) = \sum_{a,b} \int_{\lambda_{\text{AB}}} f(\mathcal{O}_a(\lambda_{\text{AB}})) \frac{\mathcal{O}_b(\lambda_{\text{AB}}) dP_{\text{AB}}(\lambda_{\text{AB}})}{p(b)} p(b).$$

Now for $p(b) > 0$, the fraction in the integrand is again a measure on $(\Omega_{\text{AB}}, \mathcal{E}_{\text{AB}})$, and Jensen's inequality gives

$$H(a|\lambda_{\text{AB}}) \leq \sum_{a,b} f\left(\int_{\lambda_{\text{AB}}} \mathcal{O}_a(\lambda_{\text{AB}}) \cdot \frac{\mathcal{O}_b(\lambda_{\text{AB}}) dP_{\text{AB}}(\lambda_{\text{AB}})}{p(b)}\right) p(b)$$

Since $p(a, b) = \int_{\lambda_{\text{AB}}} \mathcal{O}_a(\lambda_{\text{AB}}) \mathcal{O}_b(\lambda_{\text{AB}}) dP_{\text{AB}}$, the integral inside f evaluates to $p(a|b)$, so that

$$H(a|\lambda_{\text{AB}}) \leq \sum_{a,b} f(p(a|b)) p(b) = H(a|b), \quad (\text{A.2})$$

which is the data processing inequality we wanted to prove.

We now make the usual estimates known from proofs of nonnegativity of conditional mutual information or nonnegativity of Kullback-Leibler divergence [CT06, Thm. 8.6.1],

$$\begin{aligned} &H(a|\lambda_{\text{AB}}) + H(a|\lambda_{\text{AC}}) - H(a) - H(a|\lambda_{\text{AB}}\lambda_{\text{AC}}) \\ &= \sum_a \left[\int_{\lambda_{\text{AB}}, \lambda_{\text{AC}}} \mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}}) \left(-\log(\mathcal{O}_a(\lambda_{\text{AB}})) - \log(\mathcal{O}_a(\lambda_{\text{AC}})) + \log(p(a)) + \log(\mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}})) \right) dP_{\text{AB}} dP_{\text{AC}} \right] \\ &= - \sum_a \left[\int_{\lambda_{\text{AB}}, \lambda_{\text{AC}}} \mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}}) \log\left(\frac{\mathcal{O}_a(\lambda_{\text{AB}})\mathcal{O}_a(\lambda_{\text{AC}})}{p(a)\mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}})}\right) dP_{\text{AB}} dP_{\text{AC}} \right] \\ &\geq - \log \left[\sum_a \int_{\lambda_{\text{AB}}, \lambda_{\text{AC}}} \mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}}) \cdot \frac{\mathcal{O}_a(\lambda_{\text{AB}})\mathcal{O}_a(\lambda_{\text{AC}})}{p(a)\mathcal{O}_a(\lambda_{\text{AB}}, \lambda_{\text{AC}})} dP_{\text{AB}} dP_{\text{AC}} \right] = - \log \left[\sum_a \frac{p(a)p(a)}{p(a)} \right] = 0. \end{aligned}$$

Since $H(a|\lambda_{\text{AB}}\lambda_{\text{AC}})$ is defined as the integral of an a.s. nonnegative function, it is itself nonnegative, and therefore

$$H(a|\lambda_{\text{AB}}) + H(a|\lambda_{\text{AC}}) \geq H(a). \quad (\text{A.3})$$

Piecing finally the two ingredients (A.2) and (A.3) together, we find

$$I(a : b) + I(a : c) = 2H(a) - H(a|b) - H(a|c) \leq 2H(a) - H(a|\lambda_{\text{AB}}) - H(a|\lambda_{\text{AC}}) \leq H(a),$$

as was to be shown.

A.5. General proof of Theorem 2.21. In the discrete-variable case, we started with the assumption that the measurements were deterministic and noticed that if a certain combination of outcomes has positive probability, then there has to be a combination of hidden variable values, each occurring with positive probability, which produces that outcome combination.

This reasoning needs to be modified in order to apply in the general case; when dealing with non-atomic probability spaces, no single hidden variable combination has positive probability. It is therefore necessary to consider combinations of *sets* of hidden variable values, which is unfortunately somewhat technical.

LEMMA. *Let $(\Omega_1, \mathcal{E}_1, P_1), \dots, (\Omega_n, \mathcal{E}_n, P_n)$ be probability spaces and let $\Omega = \prod_{i=1}^n \Omega_i$ be equipped with the product σ -algebra $\mathcal{E} = \sigma(\prod_{i=1}^n \mathcal{E}_i)$ and the product measure $P = \prod_{i=1}^n P_i$, so that (Ω, \mathcal{E}, P) is a probability space.*

Then, for a measurable function $f : \Omega \rightarrow \{0, 1\}$ with $P(f = 1) > 0$ and any $\varepsilon > 0$, there exist measurable subsets $\Xi_i \subseteq \Omega_i$, with $P_i(\Xi_i) > 0$, such that

$$P(f = 1 \mid \Xi_1 \times \dots \times \Xi_n) > 1 - \varepsilon.$$

PROOF. This lemma can be reformulated as saying that if $\Theta \subseteq \Omega$ has positive measure, then there exist $\Xi_i \subseteq \Theta$ of positive measure such that $P(\Theta \mid \prod_{i=1}^n \Xi_i) > 1 - \varepsilon$.

We start to prove this reformulation by noting that the collection of sets which are finite disjoint unions of product sets is an algebra of sets [Hal50, 33.E]. It then follows from the approximation lemma of measure theory [Hal50, 13.D] that Θ , a set of positive measure, can be δ -approximated by a set $S(\delta)$ which is a finite union of product sets, i.e. for every $\delta > 0$ we can find such $S(\delta)$ with $P(\Theta \setminus S(\delta)) < \delta$ and $P(S(\delta) \setminus \Theta) < \delta$. We assume $\delta < P(\Theta)$, so that $P(S(\delta)) > 0$ is guaranteed.

Decomposing this $S(\delta)$ into a finite union of disjoint product sets gives

$$S(\delta) = \bigcup_{j=1}^{k(\delta)} \Xi^j(\delta)$$

for product sets $\Xi^j(\delta) = \Xi_1^j(\delta) \times \dots \times \Xi_n^j(\delta)$, which we may assume to be of positive measure (if some $\Xi^j(\delta)$ has zero measure, then it may as well be omitted). By construction, we know

$$\sum_{j=1}^{k(\delta)} P(\Xi^j(\delta) \cap \Theta) > P(\Theta) - \delta, \quad \sum_{j=1}^{k(\delta)} P(\Xi^j(\delta) \setminus \Theta) < \delta.$$

Since the second inequality states that

$$\sum_{j=1}^k \frac{P(\Xi^j(\delta) \cap \Theta)}{P(S(\delta))} \cdot \frac{P(\Xi^j(\delta) \setminus \Theta)}{P(\Xi^j(\delta) \cap \Theta)} < \frac{\delta}{P(S(\delta))},$$

and this sum is a convex combination, we conclude that there is at least one index j for which

$$\frac{P(\Xi^j(\delta) \setminus \Theta)}{P(\Xi^j(\delta) \cap \Theta)} < \frac{\delta}{P(S(\delta))} < \frac{\delta}{P(\Theta) - \delta}.$$

We define $\Xi = \prod_{i=1}^n \Xi_i$ to be equal to this $\Xi^j(\delta)$. Then

$$P(\Theta \mid \Xi) = \frac{P(\Theta \cap \Xi)}{P(\Xi \setminus \Theta) + P(\Theta \cap \Xi)} = \left(1 + \frac{P(\Xi \setminus \Theta)}{P(\Theta \cap \Xi)}\right)^{-1} > \left(1 + \frac{\delta}{P(\Theta) - \delta}\right)^{-1}.$$

For δ sufficiently small, this is $> 1 - \varepsilon$, as has been claimed. \square

We return to the main line of the proof of Theorem 2.21 and fix $\varepsilon > 0$. In a hidden variable combination like $(\ell_{AB}, \ell_{BY}, \ell_{YX}, \ell_{XA})$, each component now becomes a *set* of hidden variable values having positive probability. By the lemma, we can choose these sets in such a way that this when such a combination of hidden variables occurs, then the joint outcome is $(0, 0, 0, 0)$ with probability $> 1 - \varepsilon$. In particular, when a hidden variable combination in $(\ell_{AB}, \ell_{BY}, *, *)$ occurs, where the last two components are unspecified, then $b = 0$ with probability $> 1 - \varepsilon$. Similarly, we find a combination of sets $\kappa_{AB}, \kappa_{BY}, \kappa_{YX}, \kappa_{XA}$ producing $(1, 0, 1, 1)$ with probability $> 1 - \varepsilon$. Therefore, the combination $(\kappa_{AB}, \kappa_{BY}, \ell_{YX}, \kappa_{XA})$ yields $(1, 0, 1, 1)$ with probability $> 1 - 2\varepsilon$; it should now be clear how to complete the proof, following the steps of the discrete-variable case and bounding the probabilities in each step. Choosing ε small enough then shows that the probability to get the outcome $(0, 0, 1, 1)$ is strictly positive in contradiction with (2.16).

A.6. Separable states give rise to classical correlations. Here, we lift the restriction of finite-dimensionality from the proof of Proposition 3.17. First of all, what does separability even mean in the infinite-dimensional case? In the following, we work with arbitrary Hilbert spaces \mathcal{H} which are not necessarily separable, and put the usual trace-norm topology on $S(\mathcal{H})$; upon interpreting a quantum state on \mathcal{H} as a normal positive linear functional on $\mathcal{B}(\mathcal{H})$, this is the weak $*$ -topology. Moreover, $S(\mathcal{H})$ carries the Borel σ -algebra induced from this (metrizable) topology.

DEFINITION (cf. [HSW05]). *Let $\mathcal{H}_1, \dots, \mathcal{H}_k$ be Hilbert spaces. A state $\rho \in S(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_k)$ is separable if it lies in the closed convex hull of the set of product states.*

In general, one cannot expect a separable state to have a decomposition into a finite or infinite convex combination of product states; rather, integrals are needed [HSW05].

LEMMA. *Let $\rho \in S(\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_k)$ be separable. Then there exists a probability measure P on the set of product states such that*

$$\rho = \int_{S(\mathcal{H}_1) \otimes \dots \otimes S(\mathcal{H}_k)} (\rho_1 \otimes \dots \otimes \rho_k) dP(\rho_1 \otimes \dots \otimes \rho_k)$$

In the finite-dimensional case, one can take the measure P to have finite support, so that the integral becomes a finite convex combination.

PROOF. Since the set of product states is compact, Milman's converse to the Krein-Milman Theorem guarantess that every extreme point of the set of separable states is a product state. Then the assertion follows from Choquet's Theorem [Phe01]. \square

This should make it clear how to prove Proposition 3.17 in the general case: to a source s sending out a separable state

$$\rho_s = \int_{\prod_{\{m : sCm\}} S(\mathcal{H}_{(s,m)})} \left(\bigotimes_{\{m : sCm\}} \rho_{(s,m)} \right) dP_s \left(\bigotimes_{\{m : sCm\}} \rho_{(s,m)} \right),$$

we associate the hidden variable probability space $\Omega_s = \prod_{\{m : sCm\}} S(\mathcal{H}_{(s,m)})$ equipped with its Borel σ -algebra and its probability measure P_s , so that the hidden variable λ_s ranges over all product states $\lambda_s = \bigotimes_{\{m : sCm\}} \rho_{(s,m)}$. Concerning the conditional probabilities, (3.6) now reads

$$\mathcal{O}_m \left(\left\{ \lambda_s = \bigotimes_{\{m' : sCm'\}} \rho_{(s,m')} : sCm \right\} \right) \stackrel{\text{def}}{=} \text{tr} \left[\left(\bigotimes_{\{s : sCm\}} \rho_{(s,m)} \right) \mathcal{F}_m \right].$$

This is a continuous, and therefore measurable, function on $\Omega_s = \prod_{\{m : sCm\}} S(\mathcal{H}_{(s,m)})$.

The intuition about how this classical model works is similar to the finite-dimensional case. One may think of the hidden variable λ_s as an abstract classical description of the product state sent out by the source; each party m then receives all descriptions from all the sources it connects to, for each of these products states retains only the information concerning the system to him while throwing away the rest, and uses that information to calculate his required outcome distribution which can then get sampled in order to obtain his outcome. By construction, this produces the desired joint distribution of outcomes.

References

- [AGR81] Alain Aspect, Philippe Grangier, and Gérard Roger, *Experimental Tests of Realistic Local Theories via Bell's Theorem*, Phys. Rev. Lett. **47** (1981), 460–463. [↑4](#)
- [Ay09] Nihat Ay, *A refinement of the common cause principle*, Discrete Appl. Math. **157** (2009), no. 10, 2439–2457. [↑16](#)
- [Bar07] Jonathan Barrett, *Information processing in generalized probabilistic theories*, Phys. Rev. A **75** (2007), no. 3, 032304. [↑21](#)
- [Bel64] John S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics **1** (1964), 195–200. [↑1](#), [5](#)
- [Bel87] ———, *Speakable and unspeakable in quantum mechanics*, Cambridge University Press, Cambridge, 1987. Collected papers on quantum philosophy. [↑6](#)
- [BG11] Jonathan Barrett and Nicolas Gisin, *How Much Measurement Independence Is Needed to Demonstrate Nonlocality?*, Phys. Rev. Lett. **106** (2011), 100406. [↑15](#)
- [BGP10] Cyril Branciard, Nicolas Gisin, and Stefano Pironio, *Characterizing the Nonlocal Correlations Created via Entanglement Swapping*, Phys. Rev. Lett. **104** (2010), 170401. [↑2](#), [5](#), [6](#), [10](#), [11](#), [24](#)
- [BRGP12] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio, *Bilocal versus nonbilocal correlations in entanglement-swapping experiments*, Phys. Rev. A **85** (2012), 032119. [↑2](#), [5](#), [7](#), [8](#), [12](#), [24](#)
- [BY08] Adam Brandenburger and Noson Yanofsky, *A classification of hidden-variable properties*, Journal of Physics A: Mathematical and Theoretical **41** (2008), no. 42, 425302. [↑1](#)
- [CF12] Rafael Chaves and Tobias Fritz, *Entropic approach to local realism and noncontextuality*, Phys. Rev. A **85** (2012), 032113. [↑10](#)
- [CG07] W.N. Cottingham and D.A. Greenwood, *An Introduction to the Standard Model of Particle Physics*, Cambridge University Press, 2007. [↑4](#)
- [CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt, *Proposed Experiment to Test Local Hidden-Variable Theories*, Phys. Rev. Lett. **23** (1969), no. 15, 880–884. [↑14](#)
- [Cir80] Boris S. Cirel'son, *Quantum generalizations of Bell's inequality*, Lett. Math. Phys. **4** (1980), no. 2, 93–100. [↑10](#)
- [CL12] Bob Coecke and Raymond Lal, *Time Asymmetry of Probabilities Versus Relativistic Causal Structure: An Arrow of Time*, Phys. Rev. Lett. **108** (2012), 200403. [↑17](#)
- [Coe10] Bob Coecke, *Quantum picturalism*, Contemporary Physics **51** (2010), no. 1, 59–83. [↑9](#)
- [CR12] Roger Colbeck and Renato Renner, *Free randomness can be amplified*, Nature Physics **8** (2012), 450–454. [↑2](#), [15](#)
- [CT06] Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, Second edition, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006. [↑28](#)
- [Ebe08] Frederick Eberhardt, *Hans Reichenbach*, Stanford Encyclopedia of Philosophy, 2008. [↑15](#)
- [Eke91] Artur K. Ekert, *Quantum cryptography based on Bell's theorem*, Phys. Rev. Lett. **67** (1991), 661–663. [↑2](#)
- [EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen, *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*, Phys. Rev. **47** (1935), 777–780. [↑2](#)
- [Fin82] Arthur Fine, *Hidden Variables, Joint Probability, and the Bell Inequalities*, Phys. Rev. Lett. **48** (1982), no. 5, 291–295. [↑24](#)
- [FS12] Tobias Fritz and Robert W. Spekkens, 2012. Work in progress. [↑3](#), [4](#), [20](#)
- [GHZ90] Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger, *Bell's Theorem without inequalities*, American Journal of Physics **58** (1990), no. 12, 1131–1143. [↑18](#)
- [Haa96] Rudolf Haag, *Local quantum physics*, Second, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1996. [↑4](#)

- [Hal10] Michael J. W. Hall, *Local Deterministic Model of Singlet State Correlations Based on Relaxing Measurement Independence*, Phys. Rev. Lett. **105** (2010), 250404. [↑1](#)
- [Hal50] Paul R. Halmos, *Measure Theory*, D. Van Nostrand Company, Inc., New York, N. Y., 1950. [↑29](#)
- [Har93] Lucien Hardy, *Nonlocality for two particles without inequalities for almost all entangled states*, Phys. Rev. Lett. **71** (1993), 1665–1668. [↑18](#)
- [’H07] Gerard ’t Hooft, *On The Free-Will Postulate in Quantum Mechanics*, 2007. [quant-ph/0701097](#). [↑2](#)
- [HSW05] A. S. Holevo, M. E. Shirokov, and R. F. Werner, *Separability and Entanglement-Breaking in Infinite Dimensions*, Russian Math. Surveys **60** (2005). [↑30](#)
- [KF09] Daphne Koller and Nir Friedman, *Probabilistic graphical models*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2009. Principles and techniques. [↑5](#)
- [Jac96] Roman Jackiw, *The Unreasonable Effectiveness of Quantum Field Theory*, 1996. [arXiv:hep-th/9602122](#). [↑4](#)
- [LA09] Wolfgang Löhr and Nihat Ay, *On the generative nature of prediction*, Adv. Complex Syst. **12** (2009), no. 2, 169–194. [↑4](#)
- [McK04] Michael McKenna, *Compatibilism*, Stanford Encyclopedia of Philosophy, 2004/2009. [↑2](#)
- [PAM⁺10] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, *Random numbers certified by Bell’s theorem*, Nature **464** (2010), 1021. [arXiv:0911.3427](#). [↑2](#)
- [PBR11] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph, *On the reality of the quantum state*, 2011. [arXiv:1111.3328](#). [↑4](#)
- [Pea09] Judea Pearl, *Causality*, Second, Cambridge University Press, Cambridge, 2009. Models, reasoning, and inference. [↑3](#), [5](#)
- [Phe01] Robert R. Phelps, *Lectures on Choquet’s theorem*, Second, Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. [↑30](#)
- [Pop95] Sandu Popescu, *Bell’s Inequalities and Density Matrices: Revealing “Hidden” Nonlocality*, Phys. Rev. Lett. **74** (1995), 2619–2622. [↑4](#)
- [PR94a] Sandu Popescu and Daniel Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics **24** (1994), no. 3, 379–385. [↑18](#)
- [PR94b] ———, *Quantum nonlocality as an axiom*, Foundations of Physics **24** (1994), no. 3, 379–385. [↑18](#)
- [PV10] Károly F. Pál and Tamás Vértesi, *Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems*, Phys. Rev. A **82** (2010), 022116. [↑23](#)
- [RR56] Hans Reichenbach and Maria Reichenbach, *The Direction of Time*, Philosophy (University of California, Los Angeles), University of California Press, 1956. [↑15](#)
- [SA10] Bastian Steudel and Nihat Ay, *Information-theoretic inference of common ancestors*, 2010. [arXiv:1010.5720](#). [↑3](#), [5](#), [13](#), [15](#), [16](#)
- [Shi04] Abner Shimony, *Bell’s Theorem*, Stanford Encyclopedia of Philosophy, 2004/2009. [↑1](#), [5](#)
- [Tar51] Alfred Tarski, *A decision method for elementary algebra and geometry*, University of California Press, Berkeley and Los Angeles, Calif., 1951. 2nd ed. [↑23](#)
- [Zeh06] H. Dieter Zeh, *Quantum nonlocality vs. Einstein locality*, 2006. [http://www.rzuser.uni-heidelberg.de/~as3/nonlocality.html](#). [↑4](#)

ICFO–INSTITUT DE CIENCIES FOTONIQUES, MEDITERRANEAN TECHNOLOGY PARK, 08860 CASTELLDEFELS (BARCELONA), SPAIN

E-mail address: tobias.fritz@icfo.es